Weak-lensing-inferred scaling relations of galaxy clusters in the RCS2: mass-richness, mass-concentration, mass-bias, and more

Edo van Uitert[1,2,3], David G. Gilbank[4], Henk Hoekstra[2], Elisabetta Semboloni[2],
Michael D. Gladders[5], and Howard K. C. Yee[6]

[1] Argelander-Institut für Astronomie, Auf dem Hügel 71, 53121 Bonn, Germany
e-mail: vuitert@ucl.ac.uk
[2] Leiden Observatory, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, The Netherlands
[3] University College London, Gower Street, London WC1E 6BT, UK
[4] South African Astronomical Observatory, PO Box 9, 7935 Observatory, South Africa
[5] The Department of Astronomy and Astrophysics, and the Kavli Institute for Cosmological Physics, The University of Chicago,
5640 South Ellis Avenue, Chicago, IL 60637, USA
[6] Department of Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, Ontario, M5S 3H4, Canada

ABSTRACT

We study a sample of $\sim 10^4$ galaxy clusters in the redshift range $0.2 < z < 0.8$ with masses $M_{200} > 5 \times 10^{13}\ h_{70}^{-1}\ M_\odot$, discovered in the second Red-sequence Cluster Survey (RCS2). The depth and excellent image quality of the RCS2 enabled us to detect the cluster-mass cross-correlation up to $z \sim 0.7$. To obtain cluster masses, concentrations, and halo biases, we fit a cluster halo model simultaneously to the lensing signal and to the projected density profile of red-sequence cluster members, because the latter provides tight constraints on the cluster miscentring distribution. We parametrised the mass-richness relation as $M_{200} = A \times (N_{200}/20)^\alpha$ and find $A = (15.0 \pm 0.8) \times 10^{13}\ h_{70}^{-1}\ M_\odot$ and $\alpha = 0.73 \pm 0.07$ at low redshift ($0.2 < z < 0.35$). At intermediate redshift ($0.35 < z < 0.55$), we find a higher normalisation, which points towards a fractional increase in the richness towards lower redshift caused by the build-up of the red sequence. The miscentring distribution is well constrained. Only $\sim 30\%$ of our BCGs coincide with the peak of the dark matter distribution. The distribution of the remaining BCGs are modelled with a 2D-Gaussian, whose width increases from 0.2 to 0.4 h_{70}^{-1} Mpc towards higher masses. The ratio of width and r_{200} is constant with mass and has an average value of 0.44 ± 0.01. The mass-concentration and mass-bias relations agree fairly well with literature results at low redshift, but have a higher normalisation at higher redshifts, possibly because of selection and projection effects. The concentration of the satellite distribution decreases with mass and is correlated to the concentration of the halo.

Key words. gravitational lensing: weak – galaxies: clusters: general

1. Introduction

Observations of galaxy clusters provide a wealth of astrophysical and cosmological information. A key quantity of clusters is the mass, because it determines the relative importance of various processes such as AGN feedback. Furthermore, given a large sample of cluster masses, the cluster mass function can be determined and compared to simulations in order to constrain cosmological parameters, such as the normalisation of the matter power spectrum, σ_8, and the cosmological matter density, Ω_M (e.g. Evrard 1989; White et al. 1993); if the redshift baseline of the sample is sufficiently large, the dark energy equation of state can be constrained (e.g. Voit 2005; Allen et al. 2011).

The mass of a cluster is not a direct observable, but can be determined with a variety of techniques. The velocity distribution of cluster members has been used to derive dynamical mass estimates (e.g. van der Marel et al. 2000; Łokas et al. 2006), but these observations are generally expensive since they require spectroscopic observations of many cluster members. Additionally, assumptions on the satellite orbits are needed to convert the velocity dispersions into a mass estimate. X-rays observables can also be used to estimate the mass

(see Ettori et al. 2013, for a review), under the assumption that the hot cluster gas is in hydrostatical equilibrium. The results of Mahdavi et al. (2008, 2013) support the results from hydrodynamical simulations (e.g. Nelson et al. 2014) that clusters are generally not in hydrostatical equilibrium, which biases the X-ray-based mass estimates. Another powerful method of obtaining cluster masses is weak gravitational lensing.

In weak lensing, the distortion of the images of faint background galaxies (sources) due to the gravitational potentials of intervening structures (lenses) is measured. This signal is proportional to the excess surface mass density, which can be modelled to obtain the mass. Weak lensing does not rely on direct tracers of the potential; the distortion can be measured for any lens, out to large radii where no visible tracers can be used. Additionally, the weak lensing signal does not depend on the physical state of the matter in the clusters, and no assumptions have to be made (e.g. virial equilibrium) to measure the total projected mass. Weak lensing has been used to determine the mass of individual massive low-redshift clusters (e.g. Hoekstra 2007; Okabe et al. 2010; Hoekstra et al. 2012, 2015; Applegate et al. 2014; Gruen et al. 2013; Umetsu et al. 2014; Kettula et al. 2015),

as well as the average mass of samples of clusters and galaxy groups by stacking their lensing signals (e.g. Mandelbaum et al. 2006a; Sheldon et al. 2009; Covone et al. 2014; Ford et al. 2015).

From the weak-lensing signal, many cluster properties can be extracted, such as the cluster mass, concentration, halo bias, and miscentring distribution. The relation between these parameters can help constrain models of cluster physics. To determine the mass function, however, we need mass estimates of a large number of clusters. The lensing signal of all but the most massive clusters is generally noisy; only by stacking the signal of samples of clusters can the average mass be robustly constrained. A common solution is to determine how an observable cluster property scales with mass and can serve as a mass proxy. The Sunyaev-Zeldovich effect has been used (e.g. Williamson et al. 2011) and appears particularly useful for estimating the masses of massive clusters at high redshifts through scaling relations. Another observable property is the richness, which has the advantage that it can be determined from the same multi-colour imaging data that is used for the lensing analysis.

To determine the richness of a cluster, it is necessary to distinguish cluster galaxies from fore- and background galaxies and, if necessary, correct for contamination. Cluster members can be identified if their redshift or velocity dispersions are available, which requires either spectroscopy or observations in many bands for reliable photometric redshifts. Alternatively, cluster members can be identified using their colours as the majority of early-type galaxies in a cluster populate a narrow range in colour–magnitude space, that is, the E/S0 ridge line or the red sequence (Gladders & Yee 2000). The advantage of the latter is that observations in only two bands suffice, which makes it cheap and particularly suited for the automated detection of clusters in large imaging surveys (e.g. Gladders & Yee 2005). Additionally, only a few field galaxies reside in this regime of colour–magnitude space, which reduces the contamination.

In this paper we use optical imaging data from the second Red-sequence Cluster Survey (RCS2; Gilbank et al. 2011), both to detect clusters and determine their richness, as well as to measure their lensing signals. The survey design was chosen so as to optimise the detection of a large number of clusters using a red-sequence method (Gladders & Yee 2000). In total, $\sim 10^4$ clusters have been detected in the RCS2, spread over a wide range in optical richness, with redshifts $0.2 < z < 0.9$. In contrast, the maxBCG cluster sample (Koester et al. 2007), a catalogue of 13 823 clusters that has been detected in the Sloan Digital Sky Survey (SDSS; York et al. 2000), covers a redshift range of $0.1 < z < 0.3$, which limits its use for evolutionary studies. The redshift range of clusters in the RCS2, combined with the excellent lensing quality of the data, makes the RCS2 well suited to this purpose.

The outline is as follows. In Sect. 2, we present the various steps of the analysis: we discuss the cluster detection and richness estimates (2.1), provide details of the lensing measurement (2.2), and discuss our halo model (2.3). We highlight our novel approach of including the cluster-satellite correlation to constrain the miscentring distribution in Sect. 2.3.1. We present the mass-richness relation in Sect. 3, the cluster miscentring distribution in Sect. 4, the mass-concentration relation in Sect. 5, the satellite distribution in Sect. 6 and the mass-halo bias relation in Sect. 7. We conclude in Sect. 8. Throughout the paper we assume a WMAP7 cosmology (Komatsu et al. 2011) with $\sigma_8 = 0.8$, $\Omega_\Lambda = 0.73$, $\Omega_M = 0.27$, $\Omega_b = 0.046$ and $h = 0.7$ the dimensionless Hubble parameter. All distances quoted are in physical (rather than comoving) units unless explicitly stated otherwise.

2. Analysis

The RCS2 is a nearly 900 sq. deg imaging survey in three bands (g', r' and z') carried out with the Canada-France-Hawaii Telescope (CFHT) using the one square degree field of view camera MegaCam. The primary imaging data covers 740 sq. deg of sky, divided in 13 patches. The survey area can be split into 145 blocks of contiguous non-overlapping 2×2 deg of sky, which we use to estimate the bootstrap covariance matrices of our measurements as discussed later on. Hence in total we use 580 sq. deg. The lensing analysis is performed on the eight minute exposures in the r'-band ($r'_{\text{lim}} \sim 24.3$), which is best suited for lensing with a median seeing of 0.71''.

The photometric calibration of the RCS2 is described in detail in Gilbank et al. (2011), the lensing analysis in van Uitert et al. (2011). For details, we refer the reader to these works. In short, we measured the shapes of 2.2×10^7 galaxies using the KSB method (Kaiser et al. 1995; Luppino & Kaiser 1997; Hoekstra et al. 1998, 2000), which corresponds to a source number density of 6.3 arcmin^{-2}. Two major improvements to the lensing analysis were introduced in Cacciato et al. (2014), van Uitert et al. (2015): we used the photometric redshift catalogues from Ilbert et al. (2013) instead of the catalogues from Ilbert et al. (2009) to estimate the source redshift distribution; and secondly, we introduced a correction scheme to account for a multiplicative bias in our KSB method (due to noise bias (Kacprzak et al. 2012; Melchior & Viola 2012; Refregier et al. 2012), and galaxy blends), which affects the lensing measurement (Hoekstra et al. 2015).

2.1. Cluster detection

Galaxies clusters are identified using a modified version of the algorithm presented in Lu et al. (2009). This is a simplified version of the Gladders & Yee (2005) red-sequence cluster-finding method. The main modifications are described in Gilbank et al. (in prep.), detailing how the method was applied to RCS2 data. Briefly, the significance of a candidate cluster overdensity is determined from a count of (colour-selected) galaxies within a circle with a radius of 0.5 Mpc. Unlike the Gladders & Yee (2005) method, no magnitude weighting is applied to the galaxies, and the uncertainties in the cluster and field counts are estimated directly from Poisson statistics. Colour slices are built at regular colour intervals (which leads to irregular bins in photometric redshift), and each slice overlaps its neighbour by a quarter of the slice width. For RCS2, the method is identical except that the model colour–redshift relation was transformed to the RCS2 filters. One additional modification for RCS2 involves the centring adopted, which aims to locate the BCG via a simple two-step approach, as discussed in Appendix A.

Richnesses are estimated using an approach similar to the one outlined in Koester et al. (2007), Hansen et al. (2005, 2009), and Johnston et al. (2007) for the maxBCG cluster sample. Firstly, the number of red-sequence galaxies brighter than $M^\star + 1$ within an aperture of radius one Mpc is counted. $M^\star$ is estimated from simple stellar population synthesis models, as described in Lu et al. (2009). This count is then used to estimate r_{200} (the radius inside of which the density is 200 times the critical density, ρ_c) for each cluster using the relation $r_{200}^{\text{gal}} = 0.156 \, N_{\text{gal}}^{0.6}$ Mpc from Hansen et al. (2005). The number of red-sequence candidates brighter than $M^\star + 1$ within r_{200}^{gal} gives N_{200}, our richness estimate. Since we also apply a background subtraction, N_{200} are non-integer values.
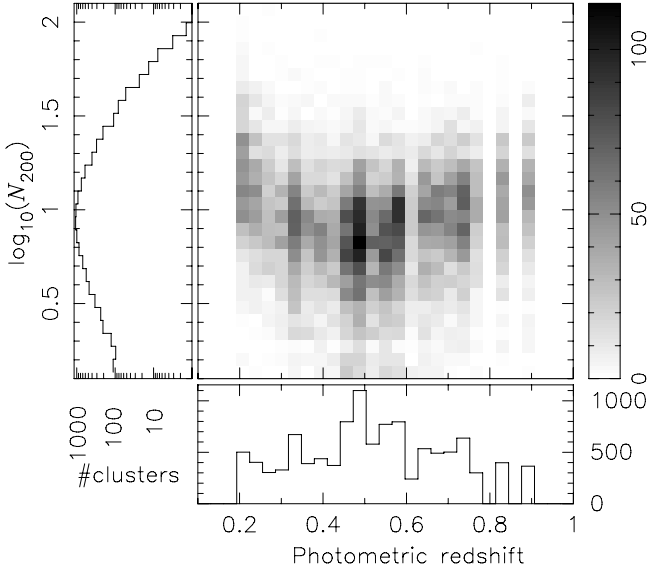
Fig. 1. Redshift versus $\log_{10}(N_{200})$, the logarithm of the number of early-type cluster members brighter than $M^\star + 1$ inside r_{200}^{gal}. The clusters cover a wide range in richness and redshift, and are therefore very well suited to studying the redshift dependence of the mass-richness relation.

Fig. 2. Comparison of our "red-sequence" redshifts to the spectroscopic redshifts from SDSS.

The maxBCG papers listed above use a variety of slightly different scaling relations which vary in both exponent and normalisation depending on precisely how r_{200} was derived. The relation we choose is similar to that given in Hansen et al. (2005), although their radius is quoted in h_{100}^{-1} Mpc and here we do not correct for the different cosmology, given the range of values in the normalisation from method-to-method. In practise, a different normalisation should just lead to counting within a different fraction of r_{200}. Therefore, mass–richness estimates within a single cluster sample should still be internally self-consistent, and to compare between samples it should suffice to apply a constant systematic scaling.

The distribution of cluster redshifts and richnesses are shown in Fig. 1. To assess the accuracy of the "red-sequence" redshifts, we match the BCGs to the galaxies with spectroscopic redshifts from the ninth data release of the SDSS (DR9; Ahn et al. 2012). In total, we find 2212 matching galaxies. We compare the redshifts in Fig. 2 and find that they agree quite well. Note that the stripes of "red-sequence" redshifts reflect the redshifts slices used to detect the clusters. Only at redshifts below $z < 0.3$ and above $z > 0.7$, the "red-sequence" redshifts are slightly overestimated. The average difference in each redshift slice is shown in Fig. 3. At low redshift, the bias is likely the result of only using red-sequence templates up to $z = 0.248$; at lower redshifts, the colours become degenerate. The cause of the bias at $z > 0.7$ is less obvious, but seems to indicate that the initial calibration of the red-sequence templates using spectroscopic data was affected by outliers (either mismatches or matches that sampled regions with poorer than average photometry). We examined these clusters in detail and decided there was no obvious reason to reject them from the analysis. We correct the cluster redshifts for this bias by fitting a third-order polynomial to these matching galaxies, and apply the same shift to all our clusters. The richnesses are recomputed using the corrected redshifts. Note that Fig. 1 shows the corrected redshifts and the corrected richnesses, the quantities used in this work.

After correcting the redshift bias, we compute the scatter and show it in the lower panel of the same figure. When we
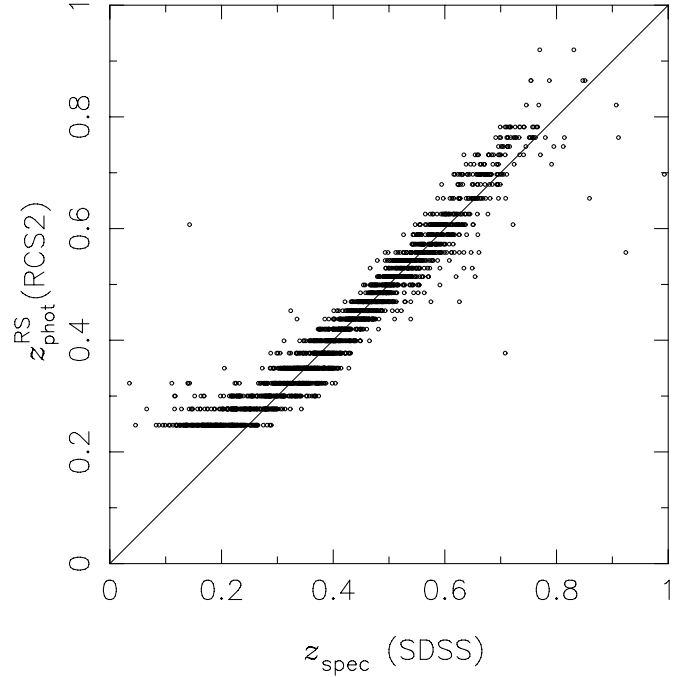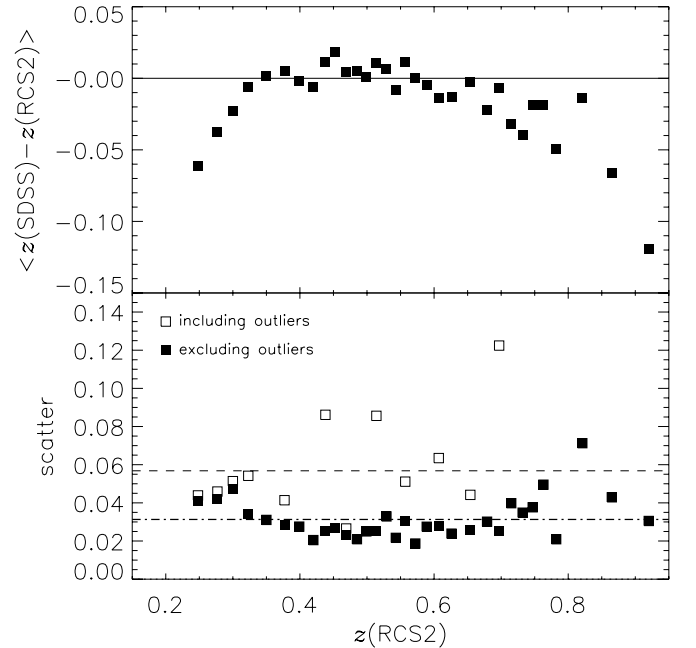

Fig. 3. Mean difference between our "red-sequence" redshifts and the spectroscopic redshifts from SDSS (*top*). Scatter between the redshifts after accounting for the bias (*bottom*). Open symbols indicate the scatter for all matches, solid ones are obtained after removing the outliers ($|z_{RCS2} - z_{SDSS}| > 0.15$). The dashed/dotted-dashed line shows the mean scatter including/excluding outliers. We correct the mean redshift bias in our analysis.

remove obvious outliers using $|z_{RCS2} - z_{SDSS}| > 0.15$, which are likely mismatches between the photometric cluster and the galaxy with spectroscopy, the scatter has a value of ~0.03 and does not vary much with redshift. Note that for "red-sequence" redshifts larger than 0.8, the bias and scatter cannot be well
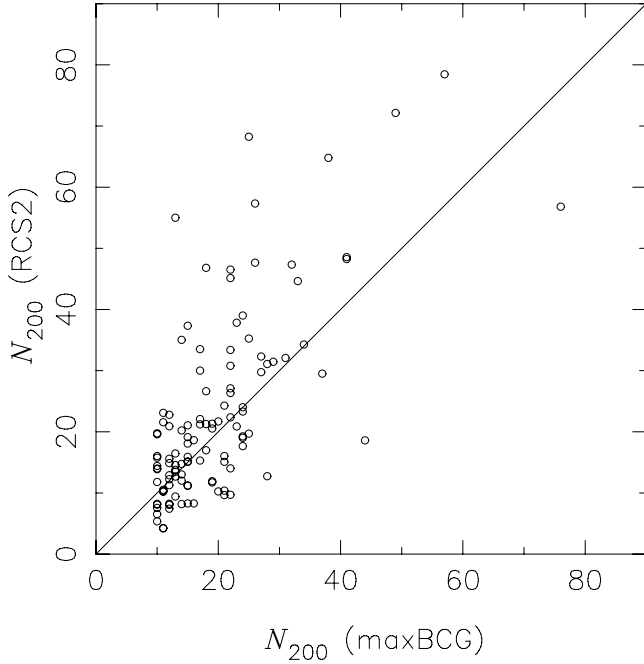
Fig. 4. Comparison of the cluster richnesses of 114 matched clusters from maxBCG and RCS2. The solid line shows the one-to-one relationship.

determined because of the low number of matches. For this work that is not important, as we restrict our analysis to clusters with $z < 0.8$.

To evaluate our richness estimates, we match our clusters to the maxBCG catalogue. We find 114 matches and compare the N_{200} estimates in Fig. 4. The richnesses of clusters in the RCS2 appear systematically larger at the high-richness end. This may partly be attributed to an improved deblending in the RCS2 because of the better observing conditions. The scatter between the estimates is quite large, part of which may be attributed to particular settings in the cluster detection algorithm that deal with mergers and projections on the sky.

Not all detections in the cluster catalogue are real clusters: a fraction of the clusters may actually correspond to a chance projection of galaxies rather than to a real cluster. These false detections have presumably a different lensing mass than the real clusters of that richness, but since both the richness and mass are affected simultaneously, the bias on the scaling relation is expected to be small. The fraction of real clusters is called the purity, which is generally a function of richness and redshift, but also depends on the cluster detection algorithm. Therefore, to determine the actual value of the purity for our cluster sample, we need to apply the detection algorithm to mock data that mimic the RCS2, which has not yet been done. The false positives do not add random noise, but a coherent (but likely lower) lensing signal. How large the impact is on the lensing mass, needs to be addressed with simulations. Note that the detection significance in our cluster finder is high, favouring a high purity over a high completeness.

Figure 1 shows that our cluster sample is incomplete at the low richness end. In principle, this should not affect our results, as the average lensing signal only becomes noisier if we miss clusters in our sample. However, to use the cluster sample to constrain cosmological parameters, a detailed knowledge of the completeness function is a prerequisite.

2.2. Lensing measurement

The shapes of source galaxies are distorted by the gravitational potentials of clusters. In the regime where the surface mass density is sufficiently small, the gravitational shear can be approximated by averaging the ellipticities of source galaxies (for a review of cluster lensing, see Hoekstra et al. 2013). We determine the tangential component of the shear in radial bins centred on the BCGs,

$$\langle \gamma_t \rangle(r) = \frac{\Delta \Sigma(r)}{\Sigma_{\mathrm{crit}}}, \tag{1}$$

which is related to the surface density contrast,

$$\Delta \Sigma(r) = \bar{\Sigma}(<r) - \bar{\Sigma}(r), \tag{2}$$

the difference between the mean projected surface density enclosed by r and the mean projected surface density at a radius r. The shear signal at small scales around massive clusters is so large that it no longer can be approximated as being linearly related to the galaxy ellipticities. We accounted for this when we computed the models (see Sect. 2.3). Σ_{crit} is the critical surface density:

$$\Sigma_{\mathrm{crit}} = \frac{c^2}{4\pi G} \frac{D_{\mathrm{s}}}{D_{\mathrm{l}} D_{\mathrm{ls}}}, \tag{3}$$

with D_{l}, D_{s} and D_{ls} the angular diameter distance to the lens, the source, and between the lens and the source, respectively (Bartelmann & Schneider 2001). Since we lack redshifts for the sources, we select galaxies with $22 < m_{r'} < 24$ that have a reliable shape estimate (ellipticities smaller than one, no SExtractor flag raised) as our source sample. To determine their approximate redshift distribution, we apply identical magnitude cuts to the publicly available photometric redshift catalogue of Ilbert et al. (2013). We use this redshift distribution to compute the mean lensing efficiency, $\langle D_{\mathrm{ls}}/D_{\mathrm{s}} \rangle$, as a function of lens redshift, accounting for the changing average weight of the sources as a function of apparent magnitude, and accounting for the impact of outliers in the catalogue. This procedure is outlined in Appendix C of Cacciato et al. (2014).

In Appendix B of van Uitert et al. (2015), we present a test of the robustness of our measurement algorithm. In short, we measure the lensing signal using different source magnitude cuts, and for each cut we recompute the lensing signal (after deriving new lensing efficiencies, random signals, noise bias corrections and source galaxy contamination corrections). Both at low and high redshifts, the resulting lensing measurements are consistent, suggesting that this method of measuring tangential shear is robust.

Here we present results from an additional test. Since we only select galaxies with a flux radius that is at least 10% larger than the local PSF size, it is possible that we systematically remove the smallest, highest redshift galaxies from our analysis; we do not account for that when we compute the lensing efficiencies, which could potentially lead to biases. We check that by removing the 10% smallest objects (in terms of r_{h}) from the photo-z catalogues of Ilbert et al. (2013), and recompute the lensing efficiencies. At low lens redshifts, the results are practically unchanged. At $z = 0.7$, the highest mean lens redshift that we study in this work, the lensing efficiencies only decrease by 3%. Removing the smallest 40% of objects (much more than we do in practice) leads to a decrease of 8%. Hence the effect is not completely negligible, but it is smaller than our statistical errors and unlikely to lead to a significant bias.
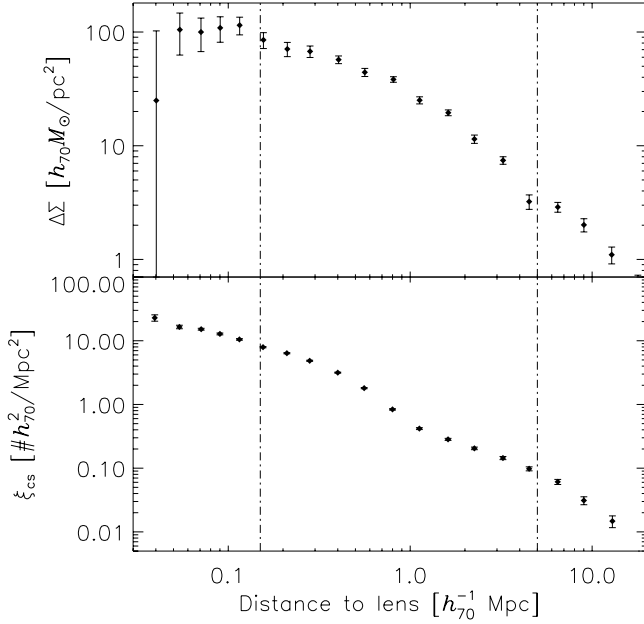
Fig. 5. Stacked lensing signal (*top*) and cluster-satellite correlation (*bottom*) measured for all clusters with $N_{200} > 2$ and $0.2 < z < 0.8$ in the RCS2. The vertical dot-dashed lines indicate the fitting range for the cluster halo model. Both measurements are shown for illustration only.

As mentioned before, the cluster redshifts have been corrected for a mean bias. This correction is uncertain for redshifts larger than 0.8 due to the low number of matches. Hence we restrict ourselves to clusters at $z < 0.8$. The scatter of the lens redshifts also affects the lensing measurement. We estimate the effect in Appendix B, and find that the impact on the lensing signal is at the per cent level. Therefore, it can be safely ignored.

The distortions induced by weak lensing are much smaller than the intrinsic ellipticities of the sources. The lensing measurement of a single cluster is therefore generally very noisy. To improve the signal-to-noise, the lensing signal is stacked for a sample of clusters that have similar properties (e.g. within a certain richness range). Stacking the lensing signal has the additional advantage that the contribution from uncorrelated structures, as well as from potential small-scale residual systematics, averages out. The lensing signal of the total cluster sample that is used in this work is shown in Fig. 5.

On large scales, residual systematics might affect the lensing signal. These systematics are commonly removed by measuring the lensing signal around random points and subtracting that from the real signal. Such a correction could also remove some real signal, hence we do not apply this correction, but use it as a test of the quality of our catalogues. The mean random signal, averaged over the 145 non-overlapping blocks of 2×2 deg each, is consistent with zero in the radial range that we use in this work.

2.2.1. Contamination

A fraction of our source galaxies is physically associated with the clusters. They are not lensed and therefore dilute the lensing signal. We cannot remove them from the source sample because we lack redshifts. We could remove the bright early-type cluster members using their colours, but the faint cluster members cannot be efficiently removed because their red sequence is not well

defined, and because many of them are blue (Hoekstra 2007). Fortunately, we can account for the dilution of the lensing signal by measuring the excess source galaxy density around the lenses, $f_{cg}(r)$, and boost the lensing signal with $1 + f_{cg}(r)$. As a further precaution, we exclude measurements on scales $<150\ h_{70}^{-1}$ kpc in our analysis.

This correction implicitly assumes that the satellite galaxies are randomly oriented. If the satellites are preferentially radially aligned to the lens, however, the contamination correction may be too low. Most recent studies in this field report no significant radial alignment (e.g. Sheldon et al. 2009; Hao et al. 2011; Schneider et al. 2013; Chisari et al. 2014; Sifón et al. 2015), although some earlier work claimed that such an alignment exists (e.g. Pereira & Kuhn 2005; Faltenbacher et al. 2007).

We do not account for the bias in the contamination correction which results from the blocking of the background sky by large (foreground) cluster galaxies (Simet & Mandelbaum 2015). In van Uitert et al. (2015), we estimate it for LRGs in the RCS2 and find that the correction is biased by 5% for low-redshift LRGs at a projected separation of 50 h_{70}^{-1} kpc. For higher redshifts, and at larger separations, the bias quickly decreases, hence it is safe to neglect it here.

Magnification by the clusters can also increase or decrease the source density, and hence bias the contamination correction. The ratio of the lensed and unlensed source counts (i.e. the bias) is given by $\mu^{\alpha-1}$, with μ the lensing magnification and α the power law index of the flux number density distribution. Using the photometric redshift catalogue of Ilbert et al. (2013), we find that $\alpha = 1.09$ at $22 < r' < 24$. To estimate the size of this bias, we assume that the mass distribution follows an NFW profile. For a cluster at $z = 0.3$ with a mass of $M_{200} = 5 \times 10^{14}\ h_{70}^{-1}\ M_\odot$ (the largest mass we probe), we find a bias of ~5% at 150 h_{70}^{-1} kpc (the smallest lens-source separation we use). For lower masses and at larger separations, the bias becomes much smaller. This bias is smaller than our statistical errors and we can safely ignore it.

2.3. Modelling of the signal

In recent years it has become common practice to model the weak-lensing signal around galaxies and galaxy clusters using the halo model (Seljak 2000; Cooray & Sheth 2002). This model provides a statistical framework of the distribution of matter in the Universe. Basically, it assumes that this distribution can be modelled as a collection of separate dark matter haloes in which galaxies and galaxy clusters are embedded. The lensing signal comes from the haloes that host the galaxy cluster and from neighbouring dark matter haloes.

The halo models we create are similar to those described in Johnston et al. (2007). Our model consists of four terms, namely the contribution of the BCG, $\Delta\Sigma_{BCG}$, the contribution of the cluster halo, $\Delta\Sigma_{NFW}$, the contribution from miscentred haloes, $\Delta\Sigma_{mis}$, and a term that accounts for the contribution from neighbouring haloes, $\Delta\Sigma_{2h}$.

The lensing signal of the BCG is sufficiently accurately modelled as a point source. We model its mass using the BCG-halo mass scaling relation from Johnston et al. (2007): $M_{BCG} = 1.334 \times 10^{12}/(1 + [M_{200}/6.717 \times 10^{13}]^{-1.38}$, in units of $h^{-1}\ M_\odot$. The contribution of the BCG to the total lensing signal is very small on scales $>150\ h_{70}^{-1}$ and we merely add it for completeness.

The central dark matter halo is described by a Navarro-Frenk-White profile (NFW; Navarro et al. 1996). The NFW density profile is given by

$$\rho(r) = \frac{\delta_c \rho_c}{(r/r_s)(1 + r/r_s)^2},$$ (4)

with δ_c the characteristic overdensity of the halo, ρ_c the critical density for closure of the universe, and $r_s = r_{200}/c_{200}$ the scale radius, with c_{200} the concentration parameter. The NFW profile is fully specified for a given set of (M_{200}, c_{200}), with M_{200} the mass inside a sphere of radius r_{200}. We calculate the tangential shear profile using the analytical expressions provided by Bartelmann (1996) and Wright & Brainerd (2000).

For a fraction of the clusters, the adopted BCG does not correspond to the actual centre of the dark matter distribution. To compute the lensing signal of a miscentred NFW profile, we first calculate the miscentred surface mass density:

$$\Sigma(r|r_{mis}) = \frac{1}{2\pi} \int_0^{2\pi} d\theta \Sigma_{NFW} \left(\sqrt{r^2 + r_{mis}^2 - 2rr_{mis}\cos(\theta)} \right)$$ (5)

with r_{mis} the offset from the peak of the dark matter. Following Johnston et al. (2007), we assume that the miscentring distribution is reasonably well described by a 2D Gaussian,

$$P(r_{mis}) = \frac{r_{mis}}{\sigma_s^2} \exp\left(-\frac{1}{2}(r_{mis}/\sigma_s)^2 \right),$$ (6)

with σ_s the width of the distribution. For a given distribution, the mean surface mass density is then given by

$$\Sigma_{mis}(r) = \int_0^\infty dr_{mis} P(r_{mis}) \Sigma_{NFW}(r|r_{mis}).$$ (7)

The model lensing signal is computed as usual using $\Delta\Sigma_{mis}(r) = \overline{\Sigma}_{mis}(<r) - \overline{\Sigma}_{mis}(r)$.

At large projected separations, neighbouring clusters significantly contribute to the lensing signal. This contribution is modelled as the two-halo term from the halo model presented in Mandelbaum et al. (2006b), $\Delta\Sigma_{2h}$. To avoid that neighbouring haloes overlap, we implement the halo-exclusion prescription as detailed in van Uitert et al. (2011). The amplitude of the two-halo term is set by the linear bias parameter, which is a free parameter in the fit. The main difference with this two-halo term compared to other commonly used descriptions (e.g. Cacciato et al. 2009; van Uitert et al. 2011; Leauthaud et al. 2012) is that it is computed using the linear power spectrum rather than the non-linear one. We note that the regime where the one-halo and two-halo terms overlap (at a few Mpc) is difficult to model in general due to uncertainties in the prescription of halo exclusion and non-linear biasing. Marginalizing over the linear bias mitigates the impact of these uncertainties on the other fit parameters. We have implemented various variations of the two-halo term and we will discuss the effect this has on our results in the forthcoming sections.

Close to massive clusters, the lensing signal is so large that the relation between the source galaxy ellipticities and the gravitational shear can no longer be approximated as being linear. To account for this, we convert the model shear to the reduced shear, $\Delta\Sigma^{red}$, which is the quantity we measure observationally (Seitz & Schneider 1997; Hoekstra et al. 2000; Applegate et al. 2014):

$$\Delta\Sigma^{red} = \frac{\Delta\Sigma^{model}}{1 - \frac{\langle\beta_s^2\rangle}{\langle\beta_s\rangle^2} \frac{\Sigma^{model}}{\Sigma_{crit}}},$$ (8)
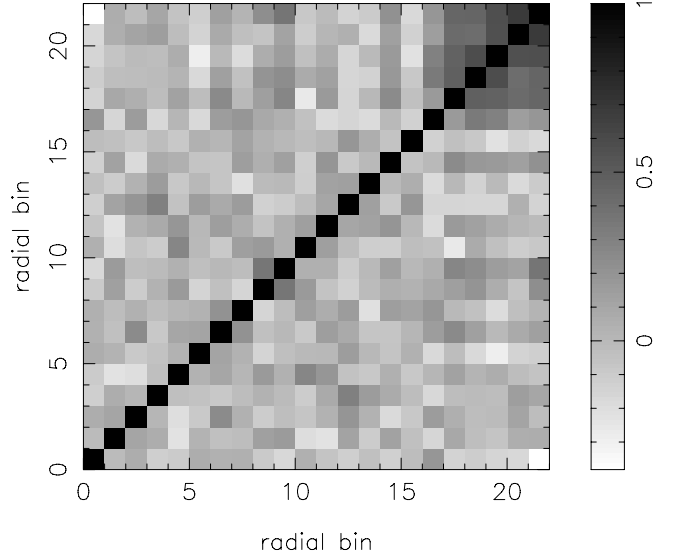
Fig. 6. Normalised covariance matrix between the radial bins of the lensing measurement and the cluster-satellite correlation for the clusters in the N4z1 sample (the fourth richness bin of the first redshift slice, see Table 2). The first 11 bins are the radial bins of the lensing measurements between $0.15 < r < 5\ h_{70}^{-1}$ Mpc, the second set of 11 bins are the radial bins of the cluster-satellite correlation between $0.15 < r < 5\ h_{70}^{-1}$ Mpc.

where

$$\beta_s = \frac{D_{ls}}{D_s} \frac{D_\infty}{D_{l\infty}},$$ (9)

with D_∞ and $D_{l\infty}$ the angular diameter distance from the observer to a source at infinity, and between the lens and a source at infinity. We determine $\langle\beta_s\rangle$ and $\langle\beta_s^2\rangle$ by integrating over the source redshift distribution, as determined from the photometric redshift catalogues of Ilbert et al. (2013). These reduced shear models are then fitted to the data. All lensing signals and model shears that we show in the following sections correspond to the reduced shear. Note that β_s and Σ_{crit} in Eq. (8) are computed using the mean lens redshift for each cluster sample, rather than integrated over the lens redshift distribution. We have checked that the difference is of the order a few percent, and can be safely ignored.

In short, the model that we fit to the lensing signal is given by

$$\Delta\Sigma_{mod}(r) = \Delta\Sigma_{BCG}(r) + p_c\Delta\Sigma_{NFW}(r)$$
$$+ (1 - p_c)\Delta\Sigma_{mis}(r) + \Delta\Sigma_{2h},$$ (10)

with p_c the fraction of clusters that is correctly centred. Rather than p_c we fit q, defined as $p_c \equiv 1/[1 + \exp(-q)]$, which has an infinite range and can therefore be assigned a Gaussian prior. A high value of q corresponds to a low miscentring fraction.

2.3.1. Constraining the miscentring distribution

Figure 6 in Johnston et al. (2007) shows that the parameters that describe the miscentring distribution are not constrained by the lensing measurements. Their results are therefore sensitive to the adopted priors. If wrong priors are used, other parameters that are degenerate with the miscentring parameters may get biased, such as the concentration. To avoid such complications, we include the cluster-satellite correlation in the fit to obtain better

constraints on the miscentring distribution, bypassing the need for using informative priors on the miscentring parameters.

The satellites of a cluster trace the dark matter distribution, although the slope of their radial distribution may differ from the slope of the projected total mass distribution (e.g. Watson et al. 2010, 2012; Budzynski et al. 2012; Tal et al. 2012). Here, we use all red-sequence candidates at the cluster redshift brighter than $M^\star + 1$ as satellites. We use this magnitude cut to ensure that our selection is homogeneous and complete over the entire redshift range of our clusters. We measure their radial distribution by correlating their positions to those of the BCGs. If a fraction of the clusters is not correctly centred, this also affects the observed distribution of satellites. The cluster-satellite correlation therefore provides additional constraints on the miscentring distribution of BCGs. Note that miscentring has a significantly smaller effect on Σ than on $\Delta\Sigma$, which is illustrated in Fig. 4 in Johnston et al. (2007). However, the signal-to-noise ratio of the cluster-satellite correlation is five to ten times larger than the one from shear, so it still provides useful additional constraints.

We measure the cluster-satellite correlation as follows:

$$\xi_{cs}(r) = \bar{n}_{rs}\left(\frac{N_{rs}(r)}{N_{rand}(r)} \times \frac{N_{rand}^{tot}}{N_{rs}^{tot}} - 1\right), \tag{11}$$

with $N_{rs}(r)$ and $N_{rand}(r)$ the number of red-sequence galaxies and random points in a radial bin with a projected radial separation r from the BCG. N_{rand}^{tot} and N_{rs}^{tot} are the total number of red-sequence galaxies and random points, respectively. The part between brackets measures the overdensity of red-sequence galaxies with respect to their average number density, $\bar{n}_{rs}$. We therefore multiply it with $\bar{n}_{rs}$ to convert it to the projected density of red-sequence galaxies in counts per Mpc^2. The signal of the total cluster sample is shown in Fig. 5.

To account for zero-point and PSF variations, we scale the number of randoms to the number of red-sequence galaxies in each field separately. Secondly, we determine the ratio of the total number of red-sequence galaxies and the total number of random points in all fields as a function of position in the camera, and correct our measurement with this ratio. The purpose of this correction is two-fold. First, it ensures that the random points have exactly the same footprint as the red-sequence galaxies. Secondly, it accounts for PSF variations across the camera. In particular, the chips in the corners of the mosaic have fewer galaxies, as the average PSF is larger than in the central chips.

To model the cluster-satellite correlation, we assume that the satellites trace the surface mass density of the dark matter, but, motivated by the results from Watson et al. (2010, 2012), Budzynski et al. (2012), Tal et al. (2012) and others, we allow the concentration of the satellites, c_{gal}, to vary. Since we do not know a priori how the number of satellites is related to the surface mass density of the model, we fit this with a nuisance parameter, following $\xi_{cs}(r) = A_{gal} \times \Sigma(r)$, with A_{gal} in units $M_\odot^{-1}$, and marginalise over it.

2.3.2. Intrinsic scatter mass-richness relation

The mass-richness relation has intrinsic scatter; therefore, the best-fit lensing mass is not equal to the mean mass of the clusters in a bin. To account for this scatter, Johnston et al. (2007) integrate their models over the probability distribution of halo masses, $P(M_{200})$, given by a log-normal distribution of M_{200} for a given N_{200}, and fit the variance in $\ln(M_{200})$. The results of Becker et al. (2007) and Evrard et al. (2008) are used as a prior on the variance, which are based on a satellite kinematics study of maxBCG clusters and simulations, respectively.

Table 1. Priors on fit parameters.

Parameter	Prior-mean	Prior-sigma
$\log_{10}(M_{200})$ [$h_{70}^{-1}\ M_\odot$]	14.15	2.00
$\ln(c_{200})$	1.39	3.00
b	2.62	4.00
$\ln(\sigma_s)$ [h_{70}^{-1} Mpc]	–0.84	0.70
q	0.0	1.2
$\sigma_{\ln M|N_{200}}$	see Table 2	0.2
$\ln(c_{gal})$	1.39	3.00
A_{gal}[$h_{70}\ M_\odot^{-1}$]	0.05	0.1

More recently, Rozo et al. (2009) studied the scatter in the mass-richness relation using X-ray measurements of the maxBCG clusters, and found $\sigma_{\ln M|N_{200}} = 0.45_{-0.18}^{+0.20}$ (95% confidence limits). This variance is smaller than the one from Becker et al. (2007); the difference was likely caused by the fact that Becker et al. (2007) did not account for the miscentring of clusters. Therefore, we use the results from Rozo et al. (2009) to account for the scatter.

Since the scatter cannot be constrained by our data, it is important to estimate the prior as accurately as possible. The scatter of the mass-richness relation from Rozo et al. (2009) is computed at a fixed richness. Our richness bins span a considerable range, which broadens the actual distribution of halo masses. To obtain a more representative value for the scatter, we take the following approach. We use a mass-richness relation to predict the mass of each cluster in a particular richness bin, and adopt that as the mean of a log-normal probability distribution with a constant $\sigma_{\ln M|N_{200}} = 0.45$. We then combine the probability distributions of all clusters in that richness bin, and fit a log-normal distribution to it. We adopt the best-fit width as the scatter and list it in Table 2. We set the prior width to 0.2, which is the error on $\sigma_{\ln M|N_{200}}$ from Rozo et al. (2009).

We use the mass-richness relation from Johnston et al. (2007), $M_{200} = 8.8 \times 10^{13}(N_{200}/20)^{1.28}\ h^{-1}\ M_\odot$. In principle, we could follow an iterative approach where we update the mass-richness relation with our findings, but given that the relation we derive is not very different, and given that the derived scatter is fairly insensitive on the adopted relation, we regard this as unnecessary.

2.3.3. Model fitting

To efficiently sample parameter space and fit the models in a reasonable amount of time, we use EMCEE (Foreman-Mackey et al. 2013), the publicly available ensemble Markov chain Monte Carlo (MCMC) sampler. The free parameters in this model are M_{200}, c_{200}, b, σ_s, q, $\sigma_{\ln M|N_{200}}$, c_{gal} and A_{gal}. For convenience, we summarise the priors in Table 1. Note that for all but one parameter, $\sigma_{\ln M|N_{200}}$, the priors are uninformative, hence the results do not depend on them. We assess how the best-fit masses depend on the prior of $\sigma_{\ln M|N_{200}}$ in the results section.

We run EMCEE with 200 walkers, starting near the best-fit model of each sample. The number of steps of each walker is set to 3000. We conservatively discard the first 500 steps as the burn-in phase. The resulting 500 000 model evaluations are used to estimate the parameter uncertainties. The fit parameters and their errors which we report in the following sections correspond to

Table 2. Cluster sample details.

Bin	N_{200}	$n_{\rm clus}$	$\langle z \rangle$	$\langle N_{200} \rangle$	$N_{200}^{\rm corr}$	$\sigma_{\ln M \mid N_{200}}$	M_{200} $[10^{13}\,h_{70}^{-1}\,M_\odot]$	c_{200}	b	$\sigma_{\rm s}$ $[h_{70}^{-1}\,{\rm Mpc}]$	$p_{\rm c}$	$c_{\rm gal}$	$A_{\rm gal}$ $[\times 10^2\,h_{70}\,M_\odot^{-1}]$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
0.20 < z < 0.35													
N1z1	2.00–7.63	534	0.30	5.2	4.0	0.57	$5.0^{+0.6}_{-0.6}$	$4.8^{+2.5}_{-1.5}$	$2.6^{+0.4}_{-0.4}$	$0.20^{+0.03}_{-0.02}$	$0.39^{+0.11}_{-0.12}$	$19.5^{+13.8}_{-6.0}$	$2.8^{+0.4}_{-0.3}$
N2z1	7.63–10.47	345	0.28	9.0	7.5	0.47	$8.9^{+0.8}_{-0.9}$	$7.5^{+3.9}_{-2.2}$	$3.1^{+0.5}_{-0.4}$	$0.24^{+0.02}_{-0.02}$	$0.21^{+0.07}_{-0.06}$	$29.3^{+15.7}_{-8.6}$	$3.2^{+0.3}_{-0.3}$
N3z1	10.47–13.65	268	0.27	11.9	10.4	0.46	$7.7^{+0.9}_{-0.9}$	$3.3^{+1.4}_{-0.9}$	$2.5^{+0.4}_{-0.4}$	$0.25^{+0.03}_{-0.03}$	$0.48^{+0.09}_{-0.09}$	$12.4^{+4.8}_{-2.9}$	$5.0^{+0.5}_{-0.4}$
N4z1	13.65–18.14	216	0.27	15.8	14.3	0.46	$11.8^{+1.3}_{-1.3}$	$6.0^{+2.1}_{-1.5}$	$3.1^{+0.5}_{-0.4}$	$0.38^{+0.03}_{-0.03}$	$0.37^{+0.05}_{-0.05}$	$19.5^{+15.2}_{-7.0}$	$4.9^{+0.5}_{-0.4}$
N5z1	18.14–26.31	197	0.25	21.8	21.2	0.47	$13.8^{+1.4}_{-1.5}$	$14.4^{+9.1}_{-4.9}$	$3.8^{+0.5}_{-0.5}$	$0.38^{+0.03}_{-0.03}$	$0.30^{+0.06}_{-0.05}$	$10.9^{+4.8}_{-3.0}$	$5.9^{+0.5}_{-0.4}$
N6z1	26.31–220.00	114	0.24	38.0	36.6	0.57	$29.3^{+3.2}_{-3.0}$	$4.0^{+1.1}_{-0.8}$	$2.8^{+0.9}_{-0.9}$	$0.42^{+0.12}_{-0.10}$	$0.59^{+0.10}_{-0.10}$	$3.7^{+0.9}_{-0.7}$	$4.6^{+0.3}_{-0.3}$
0.35 < z < 0.55													
N1z2	2.00–7.63	1648	0.47	5.0	4.0	0.59	$6.4^{+0.6}_{-0.6}$	$12.7^{+7.7}_{-4.1}$	$4.8^{+0.4}_{-0.4}$	$0.20^{+0.01}_{-0.01}$	$0.17^{+0.06}_{-0.05}$	$9.3^{+1.6}_{-1.2}$	$2.6^{+0.2}_{-0.2}$
N2z2	7.63–10.47	607	0.46	8.9	7.7	0.47	$8.7^{+1.0}_{-1.1}$	$12.7^{+10.4}_{-4.8}$	$3.5^{+0.5}_{-0.4}$	$0.28^{+0.02}_{-0.02}$	$0.26^{+0.05}_{-0.06}$	$28.6^{+12.5}_{-7.4}$	$4.4^{+0.5}_{-0.4}$
N3z2	10.47–13.65	326	0.46	11.9	10.5	0.46	$11.1^{+1.5}_{-1.4}$	$4.1^{+2.2}_{-1.4}$	$4.1^{+0.6}_{-0.6}$	$0.27^{+0.03}_{-0.03}$	$0.36^{+0.07}_{-0.08}$	$10.8^{+3.9}_{-2.4}$	$4.2^{+0.5}_{-0.4}$
N4z2	13.65–18.14	216	0.47	15.7	14.2	0.46	$17.9^{+2.3}_{-2.2}$	$9.3^{+5.5}_{-3.0}$	$4.8^{+0.9}_{-0.9}$	$0.31^{+0.03}_{-0.03}$	$0.33^{+0.07}_{-0.07}$	$10.9^{+3.4}_{-2.3}$	$3.7^{+0.4}_{-0.3}$
N5z2	18.14–26.31	126	0.46	21.2	19.5	0.47	$22.2^{+3.0}_{-3.1}$	$14.1^{+10.9}_{-5.4}$	$5.9^{+0.9}_{-0.8}$	$0.40^{+0.03}_{-0.03}$	$0.30^{+0.05}_{-0.05}$	$19.6^{+11.0}_{-6.0}$	$4.5^{+0.6}_{-0.5}$
N6z2	26.31–220.00	59	0.46	36.9	34.0	0.56	$41.0^{+6.1}_{-5.9}$	$6.5^{+4.2}_{-2.4}$	$5.1^{+1.0}_{-1.0}$	$0.41^{+0.05}_{-0.05}$	$0.32^{+0.08}_{-0.07}$	$11.9^{+4.9}_{-3.1}$	$3.8^{+0.5}_{-0.4}$
0.55 < z < 0.80													
N1z3	2.00–7.63	1170	0.64	5.1	3.3	0.58	$4.9^{+0.8}_{-0.8}$	$10.4^{+14.3}_{-5.1}$	$6.3^{+0.9}_{-0.8}$	$0.23^{+0.03}_{-0.02}$	$0.35^{+0.09}_{-0.09}$	$11.6^{+4.7}_{-2.7}$	$3.0^{+0.5}_{-0.4}$
N2z3	7.63–10.47	670	0.66	9.0	6.6	0.46	$6.7^{+1.3}_{-1.3}$	$45.9^{+95.1}_{-30.4}$	$4.3^{+0.8}_{-0.8}$	$0.28^{+0.01}_{-0.01}$	$0.23^{+0.05}_{-0.05}$	$152.5^{+104.1}_{-60.5}$	$4.8^{+1.0}_{-0.7}$
N3z3	10.47–13.65	512	0.67	11.9	9.4	0.46	$11.0^{+1.8}_{-1.7}$	$35.7^{+68.0}_{-20.7}$	$5.5^{+0.9}_{-0.8}$	$0.30^{+0.02}_{-0.02}$	$0.22^{+0.05}_{-0.05}$	$36.1^{+23.2}_{-11.4}$	$3.7^{+0.6}_{-0.5}$
N4z3	13.65–18.14	340	0.67	15.5	13.0	0.46	$13.7^{+3.0}_{-2.8}$	$3.4^{+3.3}_{-1.6}$	$5.5^{+1.2}_{-1.0}$	$0.34^{+0.02}_{-0.02}$	$0.29^{+0.05}_{-0.05}$	$51.7^{+42.3}_{-19.3}$	$3.9^{+0.9}_{-0.7}$
N5z3	18.14–26.31	190	0.67	21.3	20.1	0.47	$18.1^{+3.7}_{-3.7}$	$10.3^{+12.6}_{-5.3}$	$6.4^{+1.2}_{-1.2}$	$0.39^{+0.01}_{-0.02}$	$0.31^{+0.05}_{-0.05}$	$135.8^{+112.8}_{-67.0}$	$4.4^{+1.0}_{-0.7}$
N6z3	26.31–220.00	73	0.68	36.0	33.1	0.55	$29.2^{+7.9}_{-7.2}$	$4.7^{+6.5}_{-2.5}$	$6.7^{+1.6}_{-1.4}$	$0.33^{+0.05}_{-0.06}$	$0.44^{+0.12}_{-0.10}$	$7.6^{+3.2}_{-1.9}$	$3.5^{+0.9}_{-0.6}$

Notes. (1) richness range of the bin; (2) number of clusters; (3) median redshift; (4) mean richness; (5) mean richness corrected for Eddington bias; (6) prior-mean of the scatter in the mass-richness relation; (7) halo mass; (8) concentration; (9) halo bias; (10) miscentring width; (11) centring fraction; (12) concentration of satellite distribution; (13) amplitude of cluster-satellite correlation.

the median and the 68% confidence intervals of the marginalised posterior distributions.

To fit the models to the data, we need to estimate the covariance between the data points. We do this by measuring the signal in each of our 145 2×2 deg patches, from which we create a large number of bootstrap realisations. The covariance matrices are estimated from the different realisations. We show a representative normalised covariance matrix for one of our cluster samples in Fig. 6. This figure shows that the lensing measurements are not correlated. We therefore set the off-diagonal elements of the lensing signal, and the correlation of the lensing signal with the cluster-satellite correlation, to zero. The cluster-satellite correlation itself, however, is correlated at large scales, hence we keep those off-diagonal terms. The inverse of the resulting covariance matrix is used in the fit. To correct for the bias which is introduced when a noisy covariance matrix is inverted (Hartlap et al. 2007), we multiply the inverse covariance matrix with a correction factor $(N_{\rm sample} - N_{\rm data} - 2)/(N_{\rm sample} - 1)$, where $N_{\rm sample}$ is

the number of independent samples used in the bootstrap, in our case 145, and $N_{\rm data}$ is the number of data points, 22.

2.4. Fit results

We divide the clusters in bins of richness and redshift, as detailed in Table 2. Although inherently somewhat arbitrary, these ranges were chosen such that they enable us to reliably measure and study potential trends with richness and redshift. The stacked lensing signals are shown in Fig. 7 and the cluster-satellite correlations in Fig. 8, together with the best-fit halo models. The errors on the measurements correspond to the square root of the diagonal of the covariance matrix and indicate the 68% confidence intervals. The trends in the data, such as the kink at $\sim 1\ h_{70}^{-1}$ Mpc due to the miscentring of clusters, are well reproduced by the model. We find an average reduced chi-squared value of $\chi_{\rm red}^2 = 1.1$ (with 14 d.o.f. per bin), suggesting that the data is well modelled. When we average the residuals of the fit
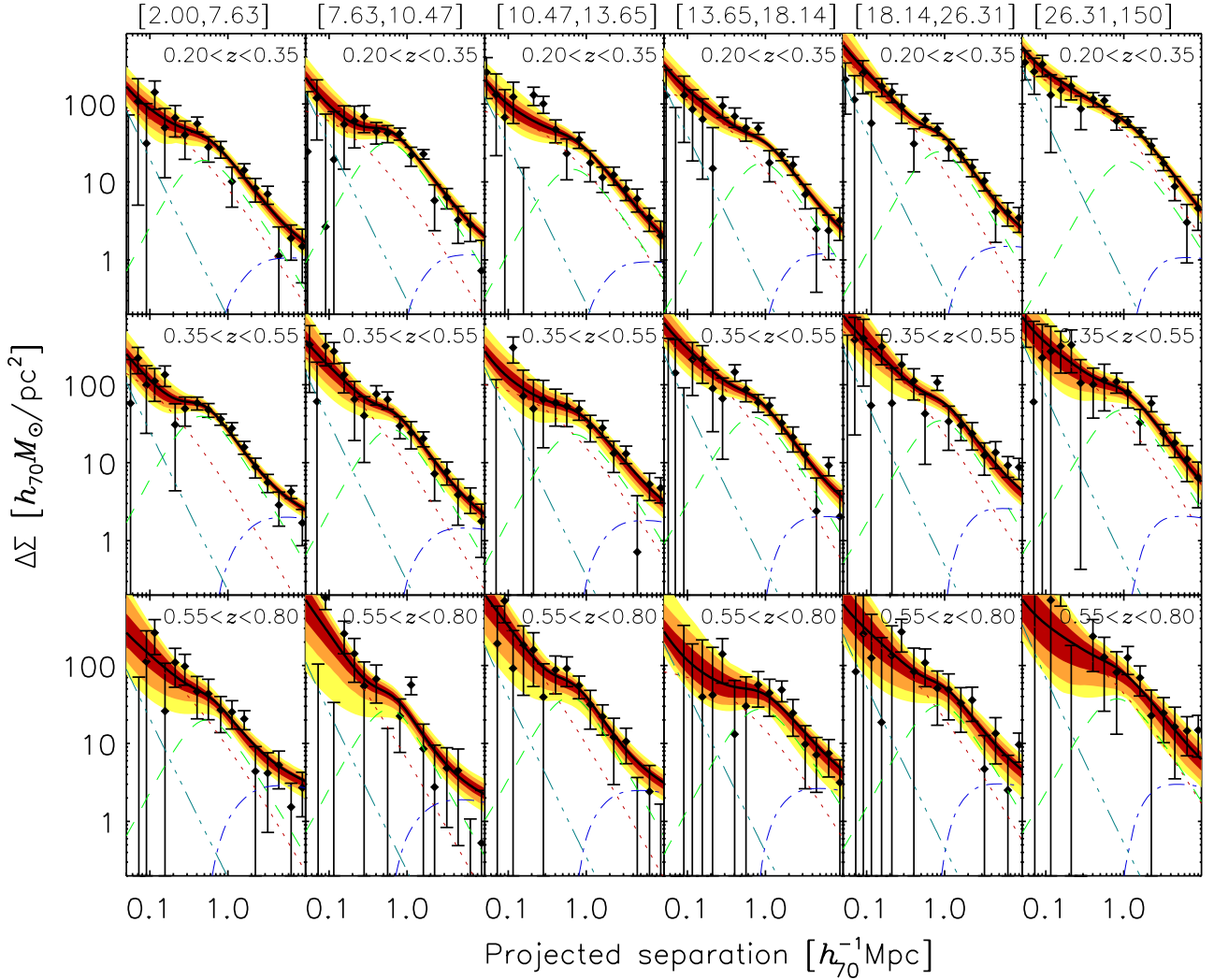
Fig. 7. Lensing signal $\Delta\Sigma$ as a function of projected separation from the BCG for the different cluster samples, split in richness (as indicated on top of each column) and redshift (indicated in each panel). The solid black lines indicate the best-fit cluster halo model, simultaneously fitted to the lensing signal and the corresponding cluster-satellite correlation signal in the range $0.15 < r < 5 \, h_{70}^{-1}$ Mpc. The dotted-dotted-dashed grey line shows the contribution from the BCG, the dotted red line the contribution from correctly centred clusters, the dashed green line the contribution from miscentred clusters, and the dotted-dashed blue line the contribution from neighbouring haloes. The dark red, orange and yellow shaded areas correspond to the 1, 2 and 3σ model uncertainty.

for all bins, we find that the models underpredict the cluster-satellite correlation at scales 1–$2 \, h_{70}^{-1}$ Mpc, exactly the regime that is difficult to model because of the overlap between the one-halo and two-halo term.

To illustrate how well the model parameters are constrained, we show the marginalised posteriors of the fit parameters for the N4z1 bin in Fig. 9, together with the priors that were used in the fit. Only $\sigma_{\ln M | N_{200}}$ cannot be constrained by the data. For the other parameters, the choice of the prior is not important as they are well constrained. A wrong choice for the prior of $\sigma_{\ln M | N_{200}}$, however, could bias our results if degeneracies exist. To investigate this, we show the two-dimensional marginalised posteriors of all pairs of parameters in Fig. 10. $\sigma_{\ln M | N_{200}}$ is only somewhat degenerate with M_{200}, but not with the other parameters. To assess the sensitivity of our results to the chosen priors, and to enable a more detailed comparison to the results of Johnston et al. (2007), we also fit our cluster halo models but *only* to the shear, adopting the priors that were used in Johnston et al. (2007). We discuss how that affects our results in the following section.

3. Mass-richness relation

In order to enable a comparison of our mass-richness relation to simulations, we have to account for Eddington bias: the observational scatter which causes clusters to preferentially move from richness ranges where the abundance of clusters is high to where it is low. This is a separate effect from intrinsic scatter, which defines the width of the halo mass distribution at a given richness if both quantities could be measured with infinite precision, which we account for in our halo model. The observational scatter is mainly caused by the field-to-field variance of the background number density of red-sequence galaxies, for which we use a global estimate. We correct N_{200} for Eddington bias using Bayes theorem. The probability distribution of the underlying N_{200} given an observed value $N_{200}^{\rm obs}$ (the posterior) is proportional to the product of the chance of having a value of $N_{200}^{\rm obs}$ given a distribution of N_{200} (likelihood) and the probability distribution of N_{200} (prior):

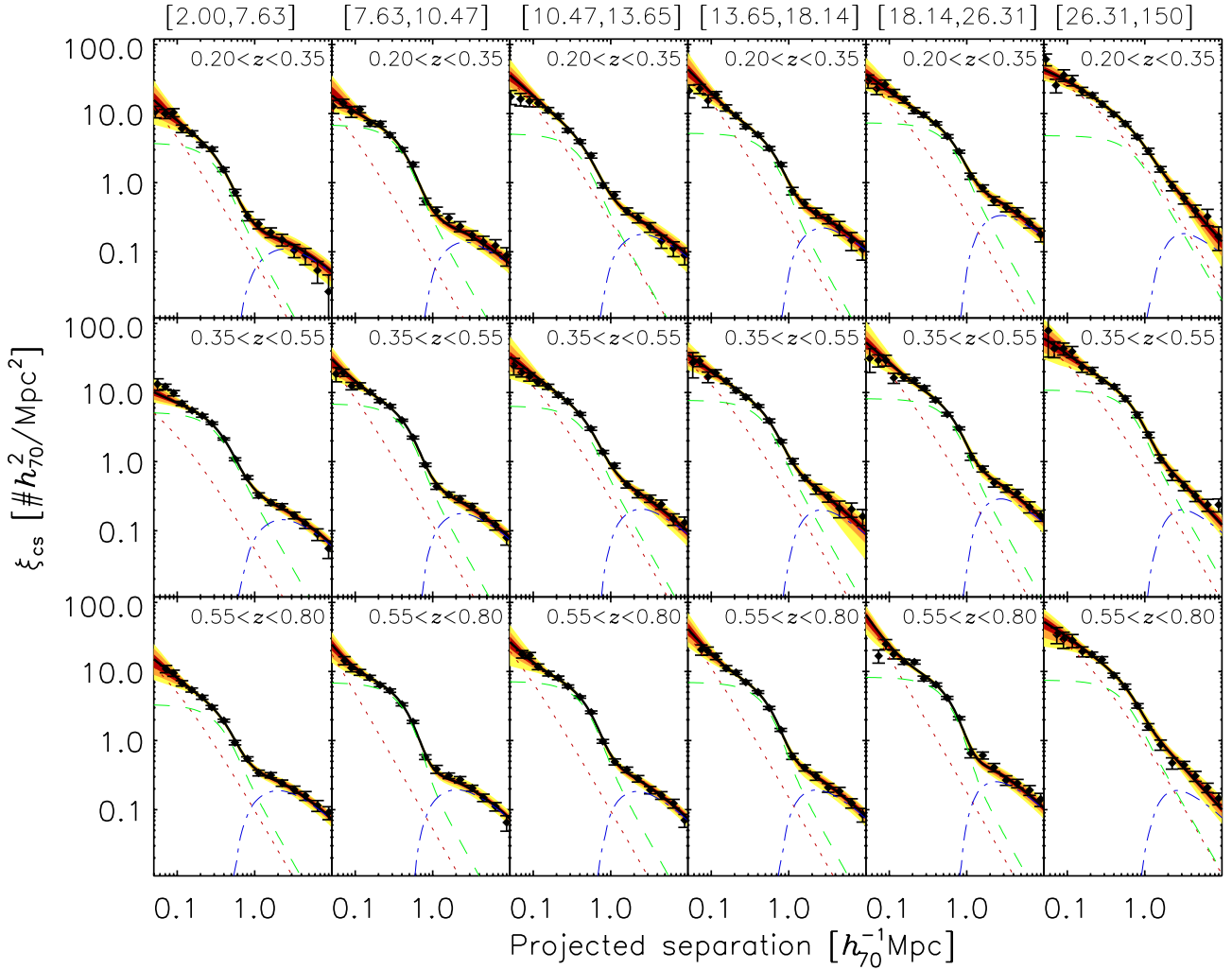$$p(N_{200}|N_{200}^{\rm obs}) \propto p(N_{200}^{\rm obs}|N_{200})p(N_{200}). \tag{12}$$

Fig. 8. Cluster-satellite correlation signal, measured using red-sequence candidates brighter than $M^\star + 1$, as a function of projected separation from the BCG, for the different cluster samples, split in richness (as indicated on top of each column) and redshift (indicated in each panel). The solid black lines indicate the best-fit cluster halo models, obtained from simultaneous fits to the lensing signal and the cluster-satellite correlation signal in the range $0.15 < r < 5\ h_{70}^{-1}$ Mpc. The dotted red line shows the contribution from correctly centred clusters, the dashed green line the contribution from miscentred clusters, and the dotted-dashed blue line the contribution from neighbouring haloes. Note that the measurements are correlated at large scale (see Fig. 5), which is accounted for in the fit. The dark red, orange and yellow shaded areas correspond to the 1, 2 and 3σ model uncertainty.

The likelihood is determined by the measurement uncertainty of N_{200}, which our cluster finder provides. The errors are ~20% larger than Poisson, but we adopt a Poisson distribution as a reasonable first order approximation for the likelihood distribution. For the prior, we could in principle use the observed richness distribution. However, as shown in Fig. 1, the cluster sample is incomplete at the low richness end by an uncertain amount, and using it as a prior would lead to an erroneous correction. Since we expect that the cluster sample is complete for approximately $N_{200} > 15$, we fit a power law to the richness distribution at $20 < N_{200} < 40$. For the prior, we replace the observed richness distribution with this power law at $N_{200} < 20$, whilst at larger richnesses we use the observed richness distribution. We sum the posteriors of all clusters in a bin, normalise it and integrate up to the mean, $N_{200}^{\rm corr}$. These values are tabulated in Table 2, as well as the uncorrected values. Only the corrected richnesses are used in the following, unless explicitly mentioned otherwise.

We show the mass-richness relation in Fig. 11. Note that the mass we show is the mean of the log-normal distribution of halo

masses we integrate over in our halo model. We fit a power-law relation of the form $M_{200} = A(N_{200}^{\rm corr}/20)^\alpha$ in each redshift slice and report the best-fit slopes and normalisations in Table 3. The errors on the amplitude are determined by marginalising over the slope, and vice versa. The likelihood contours of the fit are shown in Fig. 12. Without the correction for Eddington bias, we would have obtained $A = (14.4 \pm 0.7) \times 10^{13}\ h_{70}^{-1}\ M_\odot$ and $\alpha = 0.83 \pm 0.08$ for the $0.20 < z < 0.35$ bin, which deviates by approximately 1σ.

A number of complications limit a simple interpretation of the weak lensing mass estimates of clusters, such as intrinsic profile variations of dark matter haloes (e.g. Clowe et al. 2004; Corless & King 2007; Gruen et al. 2015) and the presence of correlated and uncorrelated structure along the line-of-sight (e.g. Metzler et al. 2001; Hoekstra 2001; Hoekstra et al. 2011). These complications mainly increase the scatter of the mass estimates, but may even lead to small (~5–10%) biases if model fitting techniques are used (Becker & Kravtsov 2011; Rasia et al. 2012). The lensing signal can be modelled in various
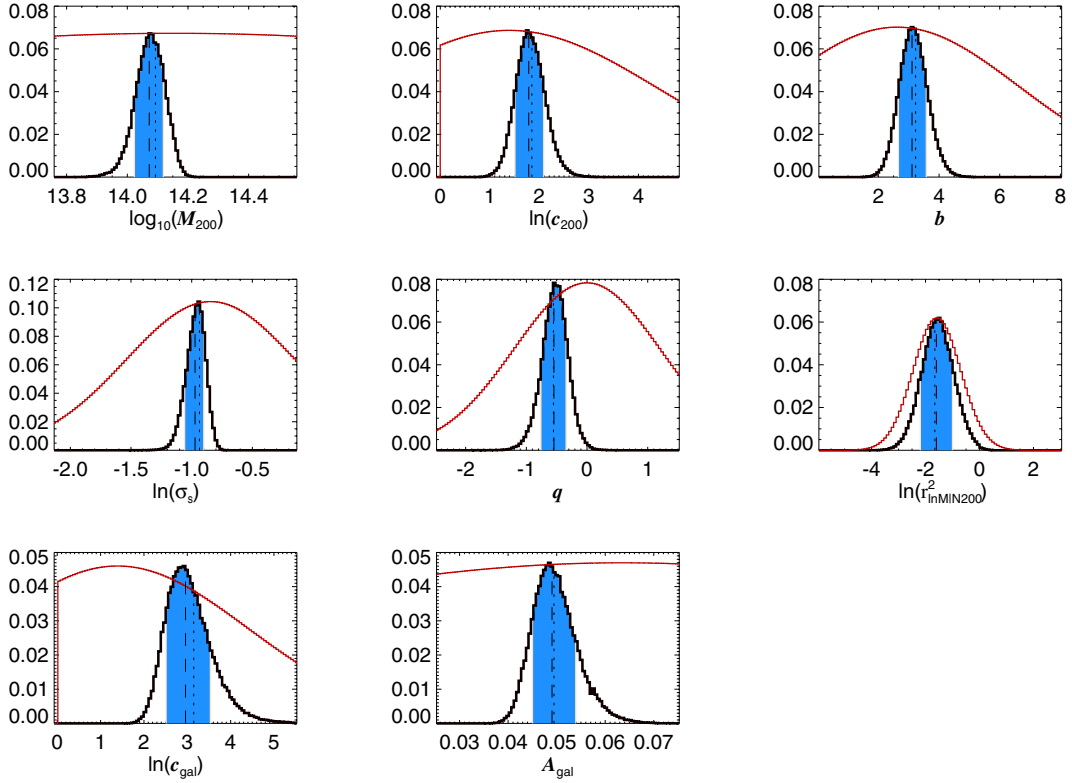
Fig. 9. Posteriors of the fitted cluster halo model parameters for the N4z1 bin, marginalised over all other parameters. Black solid lines indicate the posterior, red lines the prior. The dashed vertical line indicates the median of the marginalised posterior, the blue shaded area is the 68% confidence interval around the median, and the dotted vertical line indicates the location of the best-fit value. Dimensions as in Table 2.
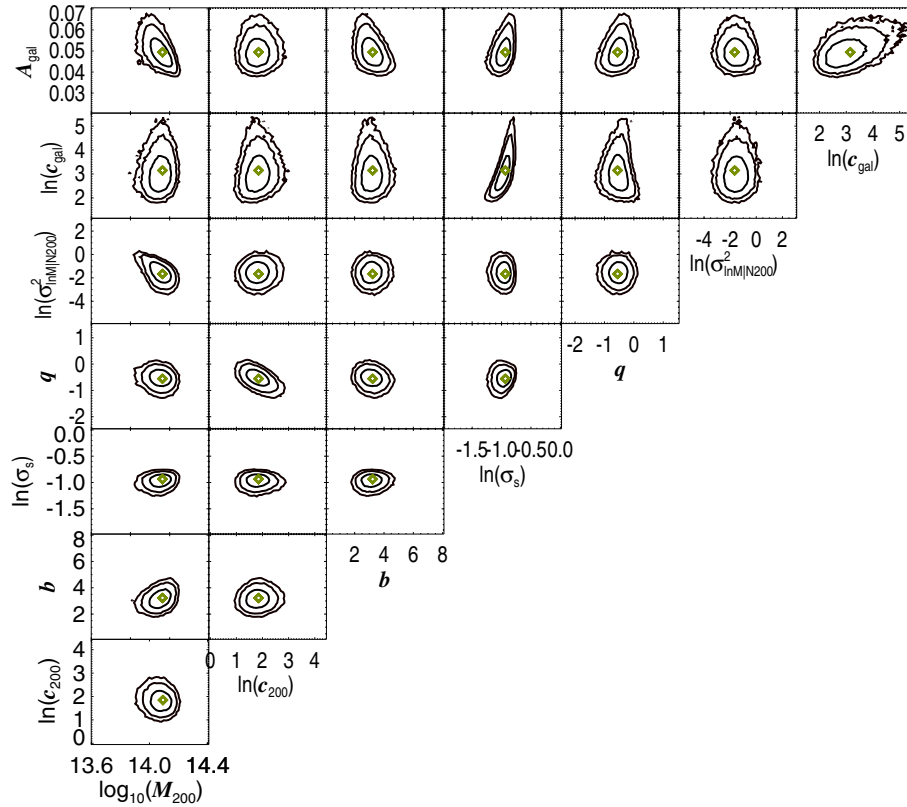

Fig. 10. Posteriors of the fitted cluster halo model parameters for the N4z1 bin, for all pairs of parameters. Shown are the 1σ, 2σ and 3σ confidence regions. The best-fit values are indicated by the green open diamonds. These plots illustrate the degeneracies that exist between the fit parameters. Dimensions as in Table 2.
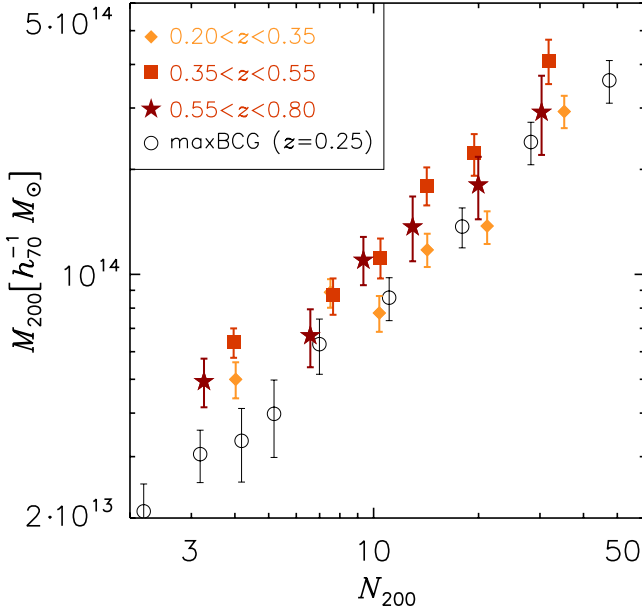
Fig. 11. Cluster mass versus richness for the cluster samples as indicated in the plot. The red circles correspond the results from the lensing analysis of the maxBCG clusters of Johnston et al. (2007), which have been boosted by a factor 1.18 to account for the impact of photometric redshift scatter (see text). All the measurements that are shown here have been corrected for Eddington bias.

Fig. 12. 67.8%, 95.4% and 99.7% confidence limits of the fits to the mass-richness relation. The black dotted lines show the results from Johnston et al. (2007) in the overlapping richness range.

Table 3. Power-law parameters of the fit between N_{200} and M_{200}.

Bin	A	α
	$[10^{13}\ h_{70}^{-1}\ M_\odot]$	
$0.20 < z < 0.35$	15.0 ± 0.8	0.73 ± 0.07
$0.35 < z < 0.55$	22.5 ± 1.7	0.86 ± 0.08
$0.55 < z < 0.80$	18.7 ± 2.2	0.78 ± 0.11
maxBCG	18.3 ± 1.1	0.98 ± 0.07
$(3 < N_{200} < 50)$		

ways, and particular choices can reduce this bias (Mandelbaum et al. 2010). More detailed numerical simulations are required to quantify this bias more precisely, e.g. as a function of mass and redshift, to interpret the results correctly. This is important for the exploitation of clusters as a reliable tool for cosmology.

3.1. Comparison to previous single-redshift results

We compare our results to the weak lensing analysis of the maxBCG cluster sample (Koester et al. 2007), a catalogue of 13 823 clusters that has been detected in the SDSS. The cluster detection algorithm employed in Koester et al. (2007) identifies the cluster red-sequence galaxies, and selects the brightest, the BCG, as centre of the cluster. The resulting cluster sample covers the richness range $10 < N_{200} < 190$ and a redshift range of $0.1 < z < 0.3$. In Sheldon et al. (2009), the cluster sample is extended to $N_{200} = 3$, which leads to a sample of ~130 000 galaxy groups and clusters. The lensing analysis of the sample is presented in Sheldon et al. (2009); the mass-richness relation is derived in Johnston et al. (2007).
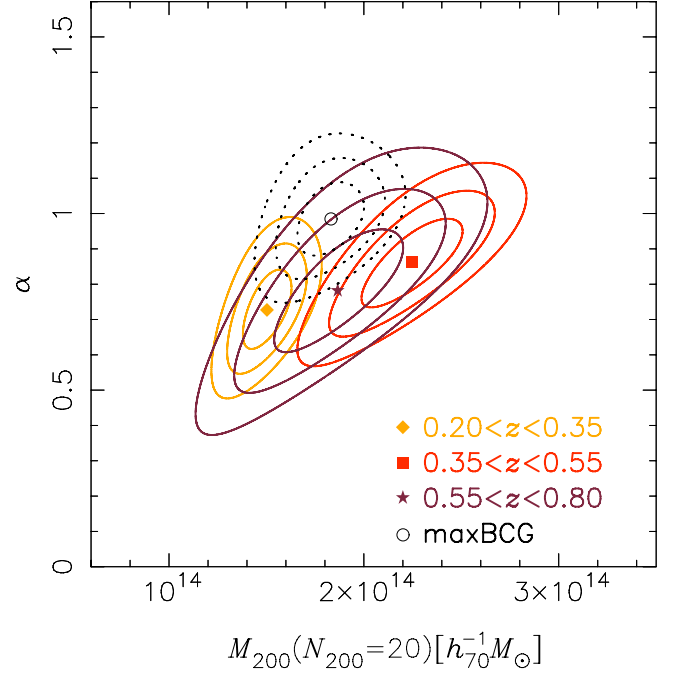
Following Mandelbaum et al. (2008b), we multiply the masses from Johnston et al. (2007) by a factor 1.18 to account for the impact of photometric redshift scatter in the lensing analysis (as was done in Rozo et al. 2010). Note that Johnston et al. (2007) use the same definition of mass as we do.

The richnesses of Johnston et al. (2007) have not been corrected for Eddington bias. We perform the correction, using a probability distribution for the maxBCG sample of $p(N_{200}) \propto (N_{200})^{-3}$ over the entire richness range, following Andreon & Hurn (2010), which is in good agreement with a slope of -3.06 that we find for our clusters with $0.2 < z < 0.35$ and $N_{200} > 15$. We adopt the mean richness as $N_{200}^{\rm obs}$ to compute the posterior, instead of stacking the posteriors of the individual clusters. This makes a negligible difference. We show the corrected results in Fig. 11. We fit the same power law in the overlapping richness range, $3 < N_{200}^{\rm corr} < 50$, and list the best-fit parameters in Table 3. The amplitude and slope of our low-redshift sample are about 3σ lower than those of the maxBCG. Figure 11 shows that this discrepancy is partly driven by the bins at $N_{200} < 10$. If we fit the relation at $N_{200} > 10$, the slopes are consistent but the amplitudes still differ by $\sim 2.5\sigma$.

There are several differences between the analyses. For instance, Fig. 4 suggests that our richnesses are somewhat larger than those of the maxBCG for rich systems; how they differ at low N_{200} is unclear as the public maxBCG catalogue only includes $N_{200} > 10$ clusters. If our richnesses are systematically larger than maxBCG for richer systems, this would tend to lower our normalisation and slope, which may partly explain the discrepancy. The number of matching RCS2 and maxBCG clusters is too low to assess this quantitatively. Also the purity of the two cluster samples may differ, as different cluster detection algorithms have been used on different data. Koester et al. (2007) show how the purity of the sample depends on particular settings of the maxBCG algorithm using mock catalogues. For maxBCG, the purity is typically of the order 90% or higher at richnesses

$N_{200} > 10$; how the purity varies at lower richnesses is not reported. Since the purity of the RCS2 cluster sample is expected to be high as well, it seems unlikely that differences in the purity could lead to differences larger than a few percent in the masses.

Also on the modelling side, there are noticeable differences. For example, we use different priors for the miscentring distribution and for the scatter between richness and halo mass. In Johnston et al. (2007), the miscentring priors are based on numerical simulations, from which a functional form is derived for q, with a corresponding p_c (the fraction of clusters that is correctly centred) ranging from 60% to 80% for richnesses in the range of $N_{200} = 10$ to $N_{200} = 100$. The width of the miscentring distribution is fixed at 0.42 h^{-1} Mpc. Our best-fit parameter suggest a p_c in the range of 20–50%, and a narrower width $\sigma_s \approx 0.2 - 0.4$ h_{70}^{-1} Mpc. It is conceivable that the miscentring distribution differs for the two catalogues, as different algorithms have been applied to different data to identify the BCG. However, if the actual miscentring distributions are similar, the different priors could cause discrepancies between the best-fit masses and concentrations due to the parameter degeneracies. The same holds for the adopted prior for the scatter in mass-richness. Motivated by this, we run our cluster halo models on the lensing data only, using the priors from Johnston et al. (2007). Note that we do not account for the differences between the two-halo terms. Nevertheless, this test will give us a reasonable impression how sensitive our results are to the adopted priors.

The resulting best-fit masses of the individual bins are consistently lower but within the 1σ error bars of the nominal results. When we refit the power-law relation, we find an amplitude and slope of $A = (12.2 \pm 1.2) \times 10^{13}$ h_{70}^{-1} $M_\odot$ and $\alpha = 0.69 \pm 0.15$ for the $0.20 < z < 0.35$ redshift slice, which is even more discrepant. Comparing this to our nominal results shows that the tighter constraints on the miscentring distribution, obtained from including the cluster-satellite correlation in the fit, also leads to smaller errors on the mass. As the amplitude is $\sim 2\sigma$ lower than our nominal result and the slope is consistent, our results do not critically depend on the adopted priors. Furthermore, we have also checked that by changing the implementation of the two-halo term, the masses do not change significantly.

The mass-richness relation has also been derived from semi-analytic galaxy formation models based on N-body simulations. For example, Hilbert & White (2010) used the Millennium Simulation (Springel 2005) and found that the derived mass-richness relation agreed well in shape and amplitude with the relation from maxBCG. Hilbert & White (2010) report an amplitude and slope of $A = (17.7 \pm 0.1) \times 10^{13}$ h_{70}^{-1} $M_\odot$ and $\alpha = 1.09 \pm 0.01$. The amplitude and slope are a bit higher than our low-redshift results. The difference is again mainly driven by the $N_{200} < 10$ results, but additionally, some difference may be caused by the different definitions of richness. Angulo et al. (2012) measure the relation between optical richness and mass at $z = 0.25$ in the Millennium-XXL simulation, which extends the Millennium and Millennium-II Simulations (Springel 2005; Boylan-Kolchin et al. 2009). A power-law slope of 1.07 is reported, steeper than our results, but again the difference may originate from the low-richness end and from differences between the richness estimators.

In Ford et al. (2015), a sample of 18 000 optically selected clusters at $0.2 < z < 0.9$ are studied in the Canada-France-Hawai Telescope Lensing Survey (CFHTLenS; Heymans et al. 2012). Clusters are identified using the 3D-Matched-Filter finder of Milkeraitis et al. (2010). Richnesses are defined as all cluster members within r_{200} that are brighter than an absolute i-band

magnitude of -19.35, as detailed in Ford et al. (2014). It is not clear how that relates to our richness estimate. Using the same parametrisation of the mass-richness relation as we do, they find an amplitude and slope of $2.7^{+0.5}_{-0.4} \times 10^{13}$ $M_\odot$ and 1.4 ± 0.1, respectively. No correction for the Eddington bias is performed, leading to a larger slope. However, the normalisation is significantly lower than what we find. The difference may be partly attributed to the definition of richness (see the discussion in Sect. 5.3 of Ford et al. 2015).

3.2. Redshift evolution

At any given richness, clusters at $0.35 < z < 0.55$ have larger masses than those at $0.2 < z < 0.35$. Whether this trend continues towards higher redshifts is not clear given the large errors on the mass for the $0.55 < z < 0.8$ clusters, although our results suggest that it does not.

Both the halo masses and the richnesses of clusters may evolve. The bulk of the change in halo mass is not expected to be physical accretion, but rather an effect called pseudo evolution (Diemer et al. 2013). Halo masses are defined with respect to a background density (usually the mean or the critical density), which evolves with redshift. Even if a halo does not accrete anything, its mass increases with time as the background density drops. If the richness of a cluster would not change, we would expect to see an increase in halo mass towards lower redshifts. The fact that we find the opposite trend and that the halo mass decreases at a given richness, points towards an evolution of the richness.

Clusters build up their red sequence over cosmic time. The cluster galaxies are stripped of their gas through tidal interactions and ram pressure stripping, which quenches their star formation (e.g. Boselli & Gavazzi 2006). Consequently, the late-type spiral galaxies that are accreted turn into early-type S0 galaxies, and subsequently appear on the E/S0 ridge line. Hence even without accreting new galaxies, N_{200}, the red-sequence richness, increases as more galaxies turn red. Various works have reported an increase of the number density of faint red-sequence cluster members toward low redshift (e.g. Loh et al. 2008; Gilbank et al. 2008; Rudnick et al. 2009; Jaffé et al. 2011; Vulcani et al. 2011). For example, Rudnick et al. (2009) measure the optical cluster luminosity function of red-sequence galaxies at $z < 0.8$. They find that at magnitudes brighter than $M^\star$, the luminosity function does not evolve much, suggesting that these cluster members are already in place. However, at fainter magnitudes, the luminosity function strongly increases towards lower redshift. This supports the view that cluster richnesses (defined with a $M^\star + 1$ magnitude limit, as we do) may become larger towards lower redshift, in line with what our results suggest. There may be other processes that could cause an evolution of N_{200}. Mass segregation could lead to more bright cluster members within the inner one Mpc, which would boost N_{gal}, leading to higher values for r_{200}^{gal} and N_{200}. Mergers of cluster members both brighter than $M^\star + 1$ would lower N_{200}, but mergers of faint red-sequence members could increase it. The build-up of the red sequence, however, is likely the dominant effect.

The redshift dependence of the mass-richness relation was also measured in Sheldon et al. (2009) for the maxBCG clusters, but due to the limited redshift range of that sample no change with redshift was found. However, in a study of the relation between X-ray luminosity and richness for the maxBCG clusters, Rykoff et al. (2008) found that the X-ray luminosity at $z = 0.28$
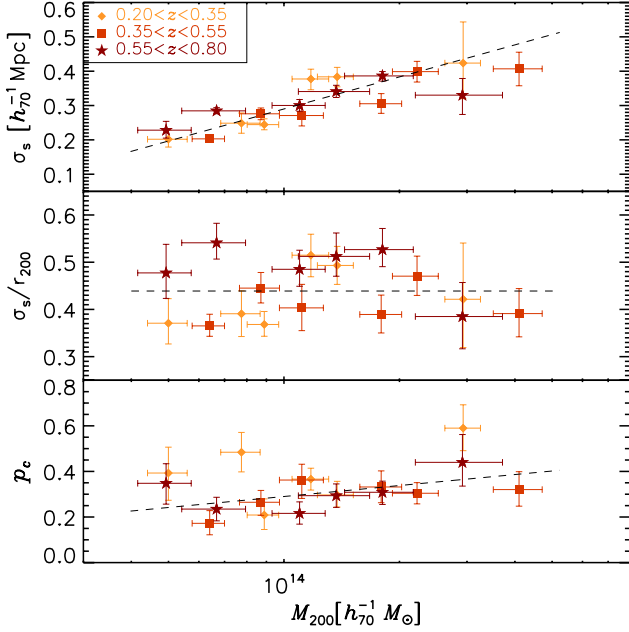
Fig. 13. Cluster mass versus the width of the miscentring distribution (*top*), versus the width divided by r_{200} (*middle*) and versus the fraction of "correctly centred" BCGs, the ones that are located at the centre of the halo (*bottom*). The black dashed lines show the fitted relation between miscentring parameter and mass, as described in the text.

is twice as high as the X-ray luminosity at $z = 0.14$. Becker et al. (2007) studied the relation between velocity dispersion and richness for the same clusters, and found that the clusters at high redshifts systematically have higher velocity dispersions. Both Becker et al. (2007) and Rykoff et al. (2008) expected the main cause to be the evolution of the N_{200} richness measure, implying a fractional decrease in N_{200} of 30–40% from $z = 0.14$ to $z = 0.28$. We ignored this effect when we compared our results in Fig. 11, but since the mean redshift of maxBCG clusters is 0.25, very close to the mean redshift of our low-redshift clusters, it is not important. If we ignore the potential evolution in halo mass, our results suggest a fractional decrease of $34 \pm 4\%$ in N_{200} from $z = 0.27$ to $z = 0.46$.

Andreon & Congdon (2014) measure the richness and mass for a sample of 23 very massive clusters with $0.15 < z < 0.55$ within a fixed aperture of 0.5 Mpc. This "aperture" mass-richness relation does not evolve with redshift. Most of the accretion and quenching of new cluster members happens at the outskirts of galaxy clusters (e.g. van der Burg et al. 2015), and may be missed when using an 0.5 Mpc aperture. Furthermore, the selection of red cluster galaxies and the computation of richness differ from our work, which may contribute to the apparent discrepancy between the results.

4. Miscentring distribution

By fitting the shear measurements together with the cluster-satellite correlation, we obtain, for the first time, tight constraints on the miscentring distribution of BCGs with respect to the centre of their dark matter haloes. Figure 13 shows that ∼30% of our BCGs are located at the centre of the halo. The distribution of the miscentred BCGs is described by a 2D-Gaussian with a standard deviation that increases from ∼0.2 to ∼0.4 h_{70}^{-1} Mpc going from

our poorest to our richest clusters; the ratio of width and r_{200} is flat with mass. Since the miscentring parameters do not appear to evolve with redshift, we combine all our cluster samples and parametrise the relations as $p_{\rm c} = A_{\rm p} + B_{\rm p} \times (\log_{10}(M_{200}) - 14)$, obtaining $A_{\rm p} = 0.29 \pm 0.02$ and $B_{\rm p} = 0.16 \pm 0.07$, and $\sigma_{\rm s} = A_{\sigma} + B_{\sigma} \times (\log_{10}(M_{200}) - 14)$, finding $A_{\sigma} = 0.29 \pm 0.01$ [h_{70}^{-1} Mpc] and $B_{\sigma} = 0.31 \pm 0.03$ [h_{70}^{-1} Mpc/$\log_{10}(h_{70}^{-1} M_{\odot})$]. Since cluster members are spread over a larger volume in more massive clusters, we expect that the mass dependency is reduced when we consider the ratio $\sigma_{\rm s}/r_{200}$. Indeed we find this ratio is consistent with a constant, for which we obtain a value of 0.44 ± 0.01 for all samples combined. The average for the low-, intermediate- and high-redshift slices are 0.42 ± 0.02, 0.40 ± 0.02 and 0.51 ± 0.02, respectively. Note that we account for the errors on both the dependent and independent parameters in the fit. The high-redshift slice has an intrinsically broader miscentring distribution, which may reflect that more clusters are still undergoing mergers at that redshift, or that the BCG is more often misidentified due to increased photometric errors. This miscentring distribution is different from the one adopted in Johnston et al. (2007), which was based on mock catalogues determined from simulations that were fit to a range of observations. There, a much higher fraction of 60–80% of the BCGs was found to be correctly centred. The width of the miscentring distribution was found to be 0.42 h^{-1} Mpc, larger than what we find. The lensing signal itself cannot discriminate between their miscentring distribution and ours, but the cluster-satellite correlation is able to break the degeneracy.

It is possible that the difference between the miscentring distributions is the result of the different algorithms used to identify the BCG. This could be tested by applying both algorithms to the same set of simulations or data. However, part of the discrepancy could also be due to how the simulations used in Johnston et al. (2007) are constructed. In these simulations, every dark matter halo is assumed to have a BCG at its centre. Hence misidentifying the BCG is the only reason why the centring fraction is less than 100%. However, in reality, the central BCG may actually be displaced by a small amount from the centre of the halo. This would push the centring fraction down and lower the width of the miscentring distribution, more in line with our findings. Also, some BCGs may be star forming, leading to too blue colours to be selected as red-sequence member (Bildfell et al. 2008).

The location of BCGs in clusters has been studied in various other works. Skibba et al. (2011) study mock catalogues and SDSS group catalogues and report that in 40% of their groups with a mass larger than 5×10^{13} $h^{-1} M_{\odot}$, the brightest galaxy is a satellite galaxy instead of the central galaxy. Using N-body simulations, Martel et al. (2014) find that the fraction where the brightest galaxy is not the nearest to the centre increases from ∼25% to ∼50%, with a higher miscentring fraction towards higher mass. Hoshino et al. (2015) study the distribution of LRGs in the redMaPPer clusters (Rykoff et al. 2014) and find that 20–30% of the brightest LRGs are not the central galaxy. As the central galaxy might be somewhat offset from the peak of the dark matter, these results might be in agreement with our results. Remarkably, both Skibba et al. (2011) and Martel et al. (2014) find that the centring fraction actually decreases with increasing mass, in contrast to what is assumed in Johnston et al. (2007). This is attributed in Martel et al. (2014) to cluster mergers, which have occurred most recently in more massive clusters as they are the last ones to form. Our centring fractions do not show a clear trend with mass.
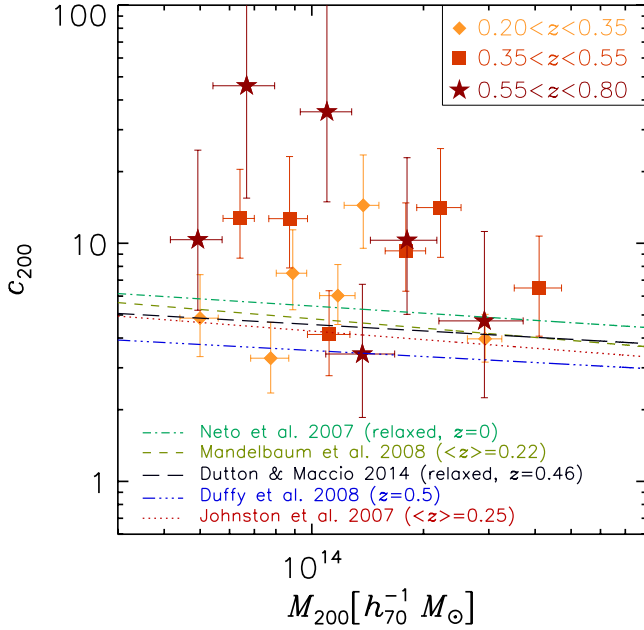
Fig. 14. Cluster mass versus concentration for the different redshift slices. The lines show the mass-concentration relations from Johnston et al. (2007), Mandelbaum et al. (2008a), Neto et al. (2007), Duffy et al. (2008) and Dutton & Macciò (2014), as indicated in the figure.

George et al. (2012) study the miscentring distribution in 129 X-ray-selected galaxy groups using their stacked weak-lensing signal, for different group centre definitions. In their model, all centres follow a 2D-Gaussian distribution, as their data does not require a correctly-centred component. Adopting the brightest group galaxy within r_{200} as the group centre, they report $\sigma_s = 24.8 \pm 12.0$ kpc, significantly smaller than the typical values for σ_s that we find. Since the groups were X-ray selected, the sample may contain relatively more relaxed systems, whose BGGs could be closer to the centre of the dark matter distribution than for the full population of groups.

Zitrin et al. (2012) studied the miscentring distribution of BCGs in 10 000 SDSS clusters, under the assumption that light traces mass. Also in that work, a much narrower miscentring distribution is reported, with a typical width of 15 kpc. Since the BCG is usually by far the brightest galaxy in a cluster, this result may not be that surprising. Zitrin et al. (2012) do not characterise the offset distribution of the misidentified clusters, which constitute about 10% of the sample. Part of the discrepancy between the results of George et al. (2012), Zitrin et al. (2012) and our findings may be caused by the fact that we include a correctly-centred component, as our data require it. Forcing all our BCGs to follow a 2D-Gaussian would lower σ_s.

5. Mass-concentration relation

Figure 14 shows the relation between mass and concentration, together with a number of literature results: the mass-concentration relation of the maxBCG clusters from Johnston et al. (2007), at a mean redshift of 0.25; the results from Mandelbaum et al. (2008a), who derived the mass-concentration relation by combining lensing measurements for L∗-type galaxies, galaxy groups traced by LRGs and the maxBCG sample, for a mean redshift of $z = 0.22$; the relation of Neto et al. (2007), derived using the Millennium Simulation at $z = 0$; the relation from

Duffy et al. (2008) at $z = 0.46$, based on large N-body simulations using the WMAP5 cosmology (Komatsu et al. 2009); and, finally, the relation from Dutton & Macciò (2014) at $z = 0.46$, derived from N-body simulations using the Planck cosmology (Planck Collaboration XVI 2014).

The concentrations for our low-redshift slice agree well with the literature results. For example, if we fit the mass-concentration relation from Johnston et al. (2007; Mandelbaum et al. 2008a) to our results, we find a relative normalisation of 1.16 ± 0.16 (1.03 ± 0.14), consistent with unity. The concentrations for our clusters in the $0.35 < z < 0.55$ range are higher. If we fit the mass-concentration relation from Dutton & Macciò (2014; Duffy et al. 2008) at $z = 0.46$, we find a relative normalisation of 1.78 ± 0.31 (2.28 ± 0.40), 2–3σ larger than unity. For the highest redshift slice, the relative normalisation is 1.67 ± 0.54 (2.12 ± 0.69) with respect to the relation from Dutton & Macciò (2014; Duffy et al. 2008) at $z = 0.67$. Note that we have ignored the error on the mass in deriving the relative normalisations, which should not matter much as the concentration changes only very weakly with mass.

To investigate how sensitive our results are to the chosen prior, and to simultaneously fitting the cluster-satellite correlation, we look at how our results change when we only fit the lensing signal, using the priors from Johnston et al. (2007). The resulting relative normalisation of the mass-concentration relation of the $0.20 < z < 0.35$ redshift slice becomes 0.85 ± 0.08 (0.76 ± 0.07) with respect to the relation from Johnston et al. (2007; Mandelbaum et al. 2008a). This is about 2σ lower than our nominal result and therefore not significant. Figure 10 shows that q and c_{200} are anti-correlated, such that forcing q to high values (lowering the miscentring fraction) pushes c_{200} down, in line with our results. Note that the constraints become tighter when we use the priors from Johnston et al. (2007) because of the use of informative priors on q and σ_s. Finally, we have also checked that our results are robust against moderate changes in modelling the two-halo term. Only if we apply a large boost to the 2-halo term, for example by using the non-linear power spectrum instead of the linear one, the best-fit concentrations become significantly higher.

Our intermediate- and high-redshift clusters have a concentration that is higher than what is expected from dark-matter-only simulations. The question of whether clusters are over-concentrated with respect to simulations has received considerable attention in recent years (e.g. Umetsu et al. 2011; King & Mead 2011; Oguri et al. 2012; Auger et al. 2013; Foëx et al. 2014; Umetsu et al. 2014). Most of these works focus on strong lensing systems, which are expected to have higher projected concentrations than typical clusters of similar mass and redshift (Hennawi et al. 2007; Oguri et al. 2009, 2012). On the other hand, weak-lensing studies of optically selected low-redshift galaxy clusters, as well as our $0.20 < z < 0.35$ results, do not show a strong deviation from the mean mass-concentration relation predicted in ΛCDM cosmologies.

Any red-sequence cluster finder that applies a spatial filtering, like ours, will have some preference for selecting over-concentrated clusters, that is, structures elongated along the line of sight. To check whether it is possible that our results are affected by selection effects, we estimate the completeness. If our sample is complete, we cannot systematically miss the under-concentrated clusters. We compute the cumulative halo mass function at the average redshift of each of the three redshift slices and multiply that with the volume in that slice. This gives us a crude estimate of the expected number of clusters above a given mass in the RCS2. Next, we use our mass-richness relation to
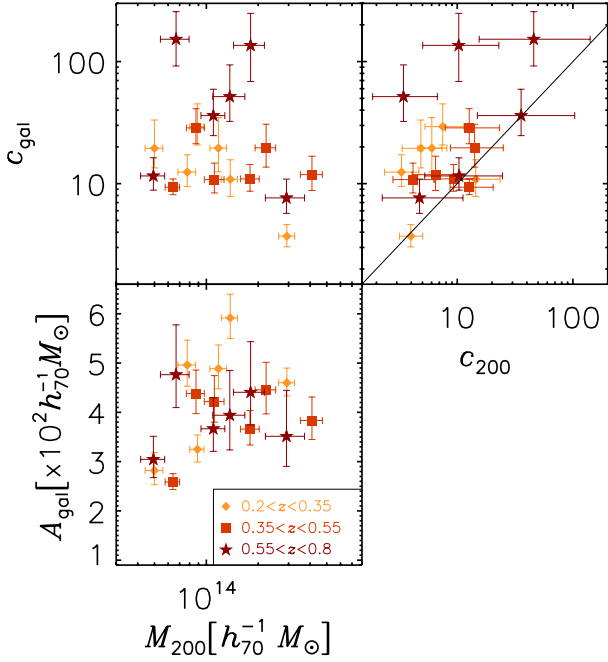
Fig. 15. Parameters describing the dark matter distribution (horizontal axis) versus parameters describing the distribution of satellites (vertical axis) for the different redshift samples. *Top left*: halo mass versus concentration of the satellite distribution; *top right*: dark matter concentration versus concentration of the satellite distribution, with the one-to-one relation as solid black line; *bottom left*: halo mass versus scaling of the surface mass density to the number density of satellites.

assign each cluster a mass and compute the observed cumulative cluster mass function. Comparing this to the theoretical prediction gives us an estimate of the completeness. For our lowest redshift slice, our cluster sample is roughly complete at $M_{200} > 10^{14}\ h_{70}^{-1}\ M_\odot$; for our intermediate-redshift clusters, we find a completeness of 70–80% in the same mass range; and for the high-redshift clusters, the completeness is less than 50%. Note that these are crude estimates, but they show that it is possible that our intermediate- and high-redshift clusters are affected by selection effects.

Another effect that might affect the concentrations are projections of pairs of clusters. Large-scale structure may be present between the clusters, which boosts the projected density at small scales. Additionally, the two parts of the projections will have lower masses and therefore on average higher concentrations, as the concentration is a decreasing function with mass, although this effect is small as the mass-concentration relation is fairly flat. If the projected separation of the two parts is small, both effects would increase the concentration. Larger projected separations are more likely, however, and would tend to lower the concentrations. Note that we find further evidence for the presence of selection and projection effects from the constraints on the halo bias, which we discuss in Sect. 7. The bias on the concentration caused by triaxiality or halo substructure is expected to be much smaller (Bahé et al. 2012).

6. Satellite distribution

We show the parameters that describe the satellite distribution in Fig. 15. c_{gal} is the most sensitive parameter to how we model the two-halo term. Given the large uncertainties in modelling the one-to-two-halo transition regime, the results need to be interpreted with great care. The reason why c_{gal} is so sensitive,

Table 4. Fit parameters of the relation between A_{gal} and M_{200}.

Bin	B [$10^{-2}\ h_{70}\ M_\odot^{-1}$]	β [$\times 10^2$]
$0.20 < z < 0.35$	4.1 ± 0.2	2.2 ± 0.6
$0.35 < z < 0.55$	3.3 ± 0.1	2.1 ± 0.5
$0.55 < z < 0.80$	3.8 ± 0.3	1.0 ± 1.0

can be seen as follows. Different implementations of the two-halo term mainly change the signal at small scales, but not the large (linear) scales. When we for example increase the signal from the two-halo term at small scales, where there is overlap with the miscentred term, the one-halo profile needs to become steeper, which is done by forcing c_{gal} upwards.

Since a different implementation mainly results in a rescaling of c_{gal}, we can only safely interpret the trends in the data, but not the values themselves. Our data suggests that the concentration of the satellite distribution decreases with mass. If we fit a linear relation between $\log_{10}(M_{200})$ and c_{gal}, we obtain a non-zero slope of -10.7 ± 2.3. Secondly, c_{gal} and c_{200} are correlated. A correlation between c_{200} and c_{gal} is expected, given that red-sequence galaxies trace the dark matter. Figure 10 suggests that this is not the result of a degeneracy between the parameters. We quantify this correlation with the Pearson coefficient, which has a value of $0.67^{+0.11}_{-0.14}$. The error on the Pearson coefficient corresponds to the 68% confidence intervals, obtained from 10 000 random realisations of the data, in which we draw new concentration values from a Gaussian whose mean equals the best-fit value and whose width corresponds to the observed error.

Finally, our measurements show that A_{gal}, the scaling between the model surface mass density and the number density of red-sequence members brighter than $M^\star + 1$, increases with mass. Since the concentration of the satellite distribution is allowed to differ from that of the dark matter, this cannot be directly interpreted as a change in scaling between the surface mass density of dark matter and the projected density of red-sequence cluster members brighter than $M^\star + 1$. A_{gal} merely serves as a nuisance parameter. For completeness, we parametrise this as $A_{gal} = B + \beta \times \log_{10}(M_{200}/10^{14})$ and list the best-fit values in Table 4. These results are not sensitive to moderate changes in the two-halo term.

7. Mass-bias relation

We show the relation between cluster mass and bias in Fig. 16, together with a number of relations from the literature, including the relation derived in Johnston et al. (2007) for the maxBCG clusters, the prediction from Sheth et al. (2001) based on an ellipsoidal collapse model and calibrated on numerical simulations, with further refinements presented in Tinker et al. (2005), and a new functional form for the mass-bias relation from Tinker et al. (2010), derived from a large set of collisionless numerical simulations based on ΛCDM. For clarity, we only show the relations at a redshift of $z = 0.27$, except for the most recent one, that is, the Tinker et al. (2010) relation; the relation with the lowest (highest) amplitude is for $z = 0.27$ ($z = 0.67$).

Our bias values increase with redshift as expected (for a compilation of redshift-dependent bias models, see Fig. 1 of Clerkin et al. 2015), but they are somewhat larger than the relations from the literature. Since our clusters do not span a wide range of mass, we only constrain the relative normalisations with respect
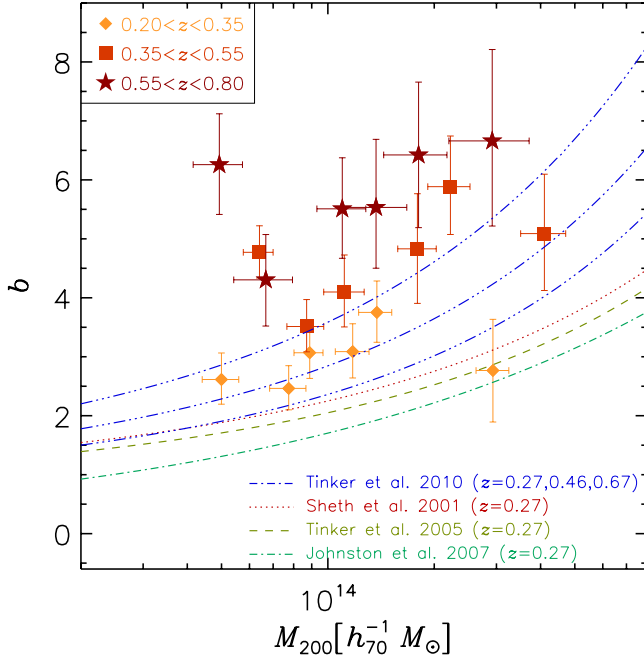
Fig. 16. Cluster mass versus bias from our halo model fits to the lensing and cluster-satellite signals for our cluster samples. The lines show the results from Tinker et al. (2010), Sheth et al. (2001), Tinker et al. (2005) and Johnston et al. (2007), as indicated in the figure.

to these relations. Fitting the relation of Tinker et al. (2010) (Johnston et al. 2007) at $z = 0.27$ to our low-redshift results, we find a relative normalisation of 1.24 ± 0.08 (1.73 ± 0.11). For the intermediate- and high-redshift slices, we find a relative normalisation of 1.45 ± 0.08 and 1.51 ± 0.11 with respect to the relation of Tinker et al. (2010) at $z = 0.46$ and $z = 0.67$, respectively. Note that we have again ignored the errors on the mass in deriving these relative normalisations, which should not matter much as the bias is only a weak function of mass.

If we use the priors from Johnston et al. (2007) and only fit the lensing signal, the errors on the bias blow up since the lensing signal at large scales is noisy. Johnston et al. (2007) fit the lensing signal out to $30~h^{-1}$ Mpc, which provides more constraining power on the amplitude of the two-halo term. Since the large-scale lensing signal is generally weak and therefore more susceptible to biases due to potentially remaining systematics in the shape catalogues, we deem that further testing of the robustness of our lensing signal is required before we can extend the fitting range. Also the cluster-satellite correlation becomes harder to measure reliably at larger scales, as it becomes increasingly sensitive to the background density.

We have tested how our results change when we implement a different prescription for the two-halo term. The biases are fairly robust against moderate changes. Only if we apply a large boost to the two-halo term, for example by replacing the linear power spectrum with the non-linear one (as is done in Cacciato et al. 2009; van Uitert et al. 2011; Leauthaud et al. 2012), the values become ~10% lower. Next to that, we note that our halo models on average underpredict the cluster-satellite correlation at 1–$2~h_{70}^{-1}$ Mpc, which may bias the halo bias high.

The difference between our bias values and the literature relations is larger for our intermediate- and high-redshift clusters. This may be caused by the same combination of effects that might explain the high concentrations. If our cluster finder preferentially selects overdense clusters, the large-scale

structure will also be preferentially oriented along the line-of-sight. Similarly, if our cluster is actually a projection, more additional structure may be present along the line-of-sight. Both would lead to a larger projected clustering of matter, and hence a larger bias.

8. Conclusion

We present the results from a combined weak-lensing and cluster-satellite correlation analysis of ~10^4 clusters in the RCS2. These clusters span a range of $0.2 < z < 0.8$ in redshift and have typical masses $M_{200} > 5 \times 10^{13}~h_{70}^{-1}~M_\odot$. We divide the clusters in three redshift slices and six richness bins, and measure the average lensing signal and cluster-satellite correlation for each sample. Satellites are identified as all red-sequence galaxies at the cluster redshift brighter than $M^\star + 1$. We model the signals simultaneously using a cluster halo model, in which we account for the miscentring distribution and the scatter between richness and mass. From these fits, we obtain the masses, the concentrations of the dark matter and of the satellite distribution, the bias, and the miscentring parameters.

We parametrise the relation between mass and richness as $M_{200} = A \times (N_{200}^{\rm corr}/20)^\alpha$ and find $A = (15.0 \pm 0.8) \times 10^{13}~h_{70}^{-1}~M_\odot$ and $\alpha = 0.73 \pm 0.07$ for our low-z clusters at $0.2 < z < 0.35$. At intermediate redshift ($0.35 < z < 0.55$), we find a higher normalisation of $A = (22.5 \pm 1.7) \times 10^{13}~h_{70}^{-1}~M_\odot$. Passive evolution and halo mass growth would lead to a higher normalisation at lower redshift, opposite to what we find. Hence we expect that this trend is driven from a fractional increase of N_{200} towards lower redshift, caused by the build-up of the red sequence. Similar trends were observed in the analyses of maxBCG clusters (Becker et al. 2007; Rykoff et al. 2008).

Our measurements provide tight constraints on the cluster miscentring distribution. Only ~30% of our BCGs are located at the centre of the halo; the remaining BCGs follow a 2D-Gaussian, whose width is approximately $0.2~h_{70}^{-1}$ Mpc at $5 \times 10^{13}~h_{70}^{-1}~M_\odot$ and increases to $0.4~h_{70}^{-1}$ Mpc for our most massive systems. The ratio of the width and r_{200} is flat with mass, with an average value of 0.44 ± 0.01. Our miscentring fraction is higher than what is typically reported in the literature, which might be caused by the commonly made assumption that the central galaxy always resides exactly at the centre of the dark matter distribution, and any miscentring is caused by not correctly identifying the central galaxy. In reality, even the central galaxy may be offset from the centre of the halo, as allowed for in our modelling, leading to smaller centring fractions.

The mass-concentration relation of our low-z cluster sample agrees well with the predictions from numerical simulations, but for our intermediate- and high-redshift clusters, a higher normalisation is preferred. We hypothesise that this is the result of two effects: a selection effect, as our cluster finder preferably selects overdense systems; and a projection effect, where two clusters are located close to the line of sight and counted as one. We find further evidence for these effects from the constraints on the bias, which show a similar trend: fair agreement with numerical simulations at low redshift, but a preference for higher values at higher redshifts.

The concentration of the satellite distribution decreases towards higher mass. It is correlated with the concentration of the dark matter. The corresponding Pearson coefficient has a value of $0.67^{+0.11}_{-0.14}$.

We have tested the robustness of our results against various changes in our halo model, such as including different implementations of the two-halo terms and the use of different priors

in the fit. These tests show that all but one parameter are robustly extracted. Only c_{gal} changes considerably when the two-halo term is modelled differently, and hence their values need to be interpreted with care. Since a different implementation mainly leads to a rescaling, the decrease of c_{gal} with mass and the correlation between c_{gal} and c_{200} are a robust find of this work.

The next step is to repeat our analysis on mock data. This is needed for a better characterisation of our cluster sample, in terms of cluster completeness, purity, and the assignment of the cluster's centres. Without that, a robust cosmological exploitation of our measurements via the halo mass function is not possible. In addition, mocks will also be very useful to help interpret our current findings, for example whether selection and projection effects can explain the high concentration and bias values that we find, or whether there is another cause. Finally, we plan to use our cluster sample with its red-sequence members for a variety of follow-up projects, including a range of alignment studies.

References

Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2012, ApJS, 203, 21
Allen, S. W., Evrard, A. E., & Mantz, A. B. 2011, ARA&A, 49, 409
Andreon, S., & Congdon, P. 2014, A&A, 568, A23
Andreon, S., & Hurn, M. A. 2010, MNRAS, 404, 1922
Angulo, R. E., Springel, V., White, S. D. M., et al. 2012, MNRAS, 426, 2046
Applegate, D. E., von der Linden, A., Kelly, P. L., et al. 2014, MNRAS, 439, 48
Auger, M. W., Budzynski, J. M., Belokurov, V., Koposov, S. E., & McCarthy, I. G. 2013, MNRAS, 436, 503
Bahé, Y. M., McCarthy, I. G., & King, L. J. 2012, MNRAS, 2249
Bartelmann, M. 1996, A&A, 313, 697
Bartelmann, M., & Schneider, P. 2001, Phys. Rep., 340, 291
Becker, M. R., & Kravtsov, A. V. 2011, ApJ, 740, 25
Becker, M. R., McKay, T. A., Koester, B., et al. 2007, ApJ, 669, 905
Bildfell, C., Hoekstra, H., Babul, A., & Mahdavi, A. 2008, MNRAS, 389, 1637
Boselli, A., & Gavazzi, G. 2006, PASP, 118, 517
Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., & Lemson, G. 2009, MNRAS, 398, 1150
Budzynski, J. M., Koposov, S. E., McCarthy, I. G., McGee, S. L., & Belokurov, V. 2012, MNRAS, 423, 104
Cacciato, M., van den Bosch, F. C., More, S., et al. 2009, MNRAS, 394, 929
Cacciato, M., van Uitert, E., & Hoekstra, H. 2014, MNRAS, 437, 377
Chisari, N. E., Mandelbaum, R., Strauss, M. A., Huff, E. M., & Bahcall, N. A. 2014, MNRAS, 445, 726
Clerkin, L., Kirk, D., Lahav, O., Abdalla, F. B., & Gaztañaga, E. 2015, MNRAS, 448, 1389
Clowe, D., De Lucia, G., & King, L. 2004, MNRAS, 350, 1038
Cooray, A., & Sheth, R. 2002, Phys. Rep., 372, 1
Corless, V. L., & King, L. J. 2007, MNRAS, 380, 149
Covone, G., Sereno, M., Kilbinger, M., & Cardone, V. F. 2014, ApJ, 784, L25
Diemer, B., More, S., & Kravtsov, A. V. 2013, ApJ, 766, 25
Duffy, A. R., Schaye, J., Kay, S. T., & Dalla Vecchia, C. 2008, MNRAS, 390, L64

Dutton, A. A., & Macciò, A. V. 2014, MNRAS, 441, 3359
Ettori, S., Donnarumma, A., Pointecouteau, E., et al. 2013, Space Sci. Rev., 177, 119
Evrard, A. E. 1989, ApJ, 341, L71
Evrard, A. E., Bialek, J., Busha, M., et al. 2008, ApJ, 672, 122
Faltenbacher, A., Li, C., Mao, S., et al. 2007, ApJ, 662, L71
Foëx, G., Motta, V., Jullo, E., Limousin, M., & Verdugo, T. 2014, A&A, 572, A19
Ford, J., Hildebrandt, H., Van Waerbeke, L., et al. 2014, MNRAS, 439, 3755
Ford, J., Van Waerbeke, L., Milkeraitis, M., et al. 2015, MNRAS, 447, 1304
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
George, M. R., Leauthaud, A., Bundy, K., et al. 2012, ApJ, 757, 2
Gilbank, D. G., Yee, H. K. C., Ellingson, E., et al. 2008, ApJ, 673, 742
Gilbank, D. G., Gladders, M. D., Yee, H. K. C., & Hsieh, B. C. 2011, AJ, 141, 94
Gladders, M. D., & Yee, H. K. C. 2000, AJ, 120, 2148
Gladders, M. D., & Yee, H. K. C. 2005, ApJS, 157, 1
Gruen, D., Brimioulle, F., Seitz, S., et al. 2013, MNRAS, 432, 1455
Gruen, D., Seitz, S., Becker, M. R., Friedrich, O., & Mana, A. 2015, MNRAS, 449, 4264
Hansen, S. M., McKay, T. A., Wechsler, R. H., et al. 2005, ApJ, 633, 122
Hansen, S. M., Sheldon, E. S., Wechsler, R. H., & Koester, B. P. 2009, ApJ, 699, 1333
Hao, J., Kubo, J. M., Feldmann, R., et al. 2011, ApJ, 740, 39
Hartlap, J., Simon, P., & Schneider, P. 2007, A&A, 464, 399
Hennawi, J. F., Dalal, N., Bode, P., & Ostriker, J. P. 2007, ApJ, 654, 714
Heymans, C., Van Waerbeke, L., Miller, L., et al. 2012, MNRAS, 427, 146
Hilbert, S., & White, S. D. M. 2010, MNRAS, 404, 486
Hoekstra, H. 2001, A&A, 370, 743
Hoekstra, H. 2007, MNRAS, 379, 317
Hoekstra, H., Franx, M., Kuijken, K., & Squires, G. 1998, ApJ, 504, 636
Hoekstra, H., Franx, M., & Kuijken, K. 2000, ApJ, 532, 88
Hoekstra, H., Hartlap, J., Hilbert, S., & van Uitert, E. 2011, MNRAS, 412, 2095
Hoekstra, H., Mahdavi, A., Babul, A., & Bildfell, C. 2012, MNRAS, 427, 1298
Hoekstra, H., Bartelmann, M., Dahle, H., et al. 2013, Space Sci. Rev., 177, 75
Hoekstra, H., Herbonnet, R., Muzzin, A., et al. 2015, MNRAS, 449, 685
Hoshino, H., Leauthaud, A., Lackner, C., et al. 2015, MNRAS, 452, 998
Ilbert, O., Capak, P., Salvato, M., et al. 2009, ApJ, 690, 1236
Ilbert, O., McCracken, H. J., Le Fèvre, O., et al. 2013, A&A, 556, A55
Jaffé, Y. L., Aragón-Salamanca, A., De Lucia, G., et al. 2011, MNRAS, 410, 280
Johnston, D. E., Sheldon, E. S., Wechsler, R. H., et al. 2007, ArXiv e-prints [arXiv:0709.1159]
Kacprzak, T., Zuntz, J., Rowe, B., et al. 2012, MNRAS, 427, 2711
Kaiser, N., Squires, G., & Broadhurst, T. 1995, ApJ, 449, 460
Kettula, K., Giodini, S., van Uitert, E., et al. 2015, MNRAS, 451, 1460
King, L. J., & Mead, J. M. G. 2011, MNRAS, 416, 2539
Koester, B. P., McKay, T. A., Annis, J., et al. 2007, ApJ, 660, 221
Komatsu, E., Dunkley, J., Nolta, M. R., et al. 2009, ApJS, 180, 330
Komatsu, E., Smith, K. M., Dunkley, J., et al. 2011, ApJS, 192, 18
Leauthaud, A., Tinker, J., Bundy, K., et al. 2012, ApJ, 744, 159
Loh, Y.-S., Ellingson, E., Yee, H. K. C., et al. 2008, ApJ, 680, 214
Łokas, E. L., Wojtak, R., Gottlöber, S., Mamon, G. A., & Prada, F. 2006, MNRAS, 367, 1463
Lu, T., Gilbank, D. G., Balogh, M. L., & Bognat, A. 2009, MNRAS, 399, 1858
Luppino, G. A., & Kaiser, N. 1997, ApJ, 475, 20
Mahdavi, A., Hoekstra, H., Babul, A., & Henry, J. P. 2008, MNRAS, 384, 1567
Mahdavi, A., Hoekstra, H., Babul, A., et al. 2013, ApJ, 767, 116
Mandelbaum, R., Seljak, U., Cool, R. J., et al. 2006a, MNRAS, 372, 758
Mandelbaum, R., Seljak, U., Kauffmann, G., Hirata, C. M., & Brinkmann, J. 2006b, MNRAS, 368, 715
Mandelbaum, R., Seljak, U., & Hirata, C. M. 2008a, J. Cosmol. Astropart. Phys., 8, 6
Mandelbaum, R., Seljak, U., Hirata, C. M., et al. 2008b, MNRAS, 386, 781
Mandelbaum, R., Seljak, U., Baldauf, T., & Smith, R. E. 2010, MNRAS, 405, 2078
Martel, H., Robichaud, F., & Barai, P. 2014, ApJ, 786, 79
Melchior, P., & Viola, M. 2012, MNRAS, 424, 2757
Metzler, C. A., White, M., & Loken, C. 2001, ApJ, 547, 560
Milkeraitis, M., van Waerbeke, L., Heymans, C., et al. 2010, MNRAS, 406, 673
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563
Nelson, K., Lau, E. T., Nagai, D., Rudd, D. H., & Yu, L. 2014, ApJ, 782, 107
Neto, A. F., Gao, L., Bett, P., et al. 2007, MNRAS, 381, 1450
Oguri, M., Hennawi, J. F., Gladders, M. D., et al. 2009, ApJ, 699, 1038
Oguri, M., Bayliss, M. B., Dahle, H., et al. 2012, MNRAS, 420, 3213
Okabe, N., Takada, M., Umetsu, K., Futamase, T., & Smith, G. P. 2010, PASJ, 62, 811
Pereira, M. J., & Kuhn, J. R. 2005, ApJ, 627, L21

Planck Collaboration XVI. 2014, A&A, 571, A16

Rasia, E., Meneghetti, M., Martino, R., et al. 2012, New J. Phys., 14, 055018

Refregier, A., Kacprzak, T., Amara, A., Bridle, S., & Rowe, B. 2012, MNRAS, 425, 1951

Rozo, E., Rykoff, E. S., Evrard, A., et al. 2009, ApJ, 699, 768

Rozo, E., Wechsler, R. H., Rykoff, E. S., et al. 2010, ApJ, 708, 645

Rudnick, G., von der Linden, A., Pelló, R., et al. 2009, ApJ, 700, 1559

Rykoff, E. S., McKay, T. A., Becker, M. R., et al. 2008, ApJ, 675, 1106

Rykoff, E. S., Rozo, E., Busha, M. T., et al. 2014, ApJ, 785, 104

Schneider, M. D., Cole, S., Frenk, C. S., et al. 2013, MNRAS, 433, 2727

Seitz, C., & Schneider, P. 1997, A&A, 318, 687

Seljak, U. 2000, MNRAS, 318, 203

Sheldon, E. S., Johnston, D. E., Scranton, R., et al. 2009, ApJ, 703, 2217

Sheth, R. K., Mo, H. J., & Tormen, G. 2001, MNRAS, 323, 1

Sifón, C., Hoekstra, H., Cacciato, M., et al. 2015, A&A, 575, A48

Simet, M., & Mandelbaum, R. 2015, MNRAS, 449, 1259

Skibba, R. A., van den Bosch, F. C., Yang, X., et al. 2011, MNRAS, 410, 417

Springel, V. 2005, MNRAS, 364, 1105

Tal, T., Wake, D. A., & van Dokkum, P. G. 2012, ApJ, 751, L5

Tinker, J. L., Weinberg, D. H., Zheng, Z., & Zehavi, I. 2005, ApJ, 631, 41

Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, ApJ, 724, 878

Umetsu, K., Broadhurst, T., Zitrin, A., et al. 2011, ApJ, 738, 41

Umetsu, K., Medezinski, E., Nonino, M., et al. 2014, ApJ, 795, 163

van der Burg, R. F. J., Hoekstra, H., Muzzin, A., et al. 2015, A&A, 577, A19

van der Marel, R. P., Magorrian, J., Carlberg, R. G., Yee, H. K. C., & Ellingson, E. 2000, AJ, 119, 2038

van Uitert, E., Hoekstra, H., Velander, M., et al. 2011, A&A, 534, A14

van Uitert, E., Cacciato, M., Hoekstra, H., & Herbonnet, R. 2015, A&A, 579, A26

Voit, G. M. 2005, Rev. Mod. Phys., 77, 207

Vulcani, B., Poggianti, B. M., Aragón-Salamanca, A., et al. 2011, MNRAS, 412, 246

Watson, D. F., Berlind, A. A., McBride, C. K., & Masjedi, M. 2010, ApJ, 709, 115

Watson, D. F., Berlind, A. A., McBride, C. K., Hogg, D. W., & Jiang, T. 2012, ApJ, 749, 83

White, S. D. M., Efstathiou, G., & Frenk, C. S. 1993, MNRAS, 262, 1023

Williamson, R., Benson, B. A., High, F. W., et al. 2011, ApJ, 738, 139

Wright, C. O., & Brainerd, T. G. 2000, ApJ, 534, 34

York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, AJ, 120, 1579

Zitrin, A., Bartelmann, M., Umetsu, K., Oguri, M., & Broadhurst, T. 2012, MNRAS, 426, 2944
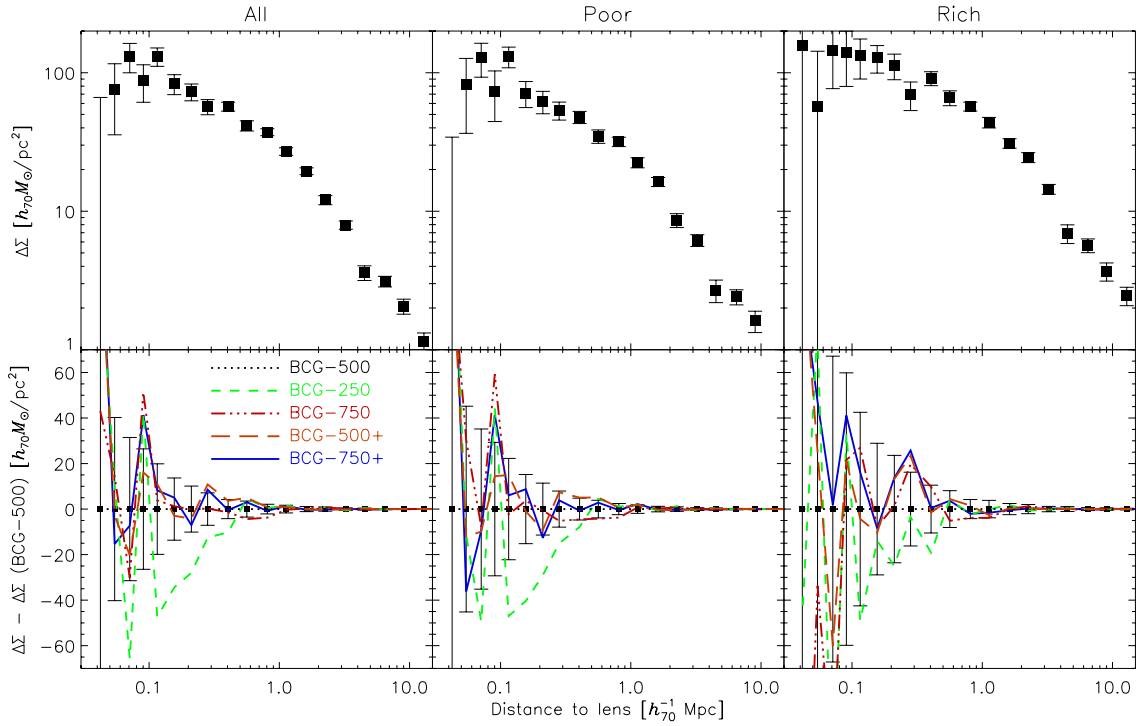
Appendix A: BCG selection

Fig. A.1. Shear signal around BCG candidates selected with the reference "BCG-500" algorithm (*top panels*). *The bottom panels* show the difference in lensing signal between various BCG selection algorithms and the reference. The *left-hand column* shows the results for all clusters, the middle one for poor clusters and the *right-hand panel* for rich clusters. The error on the difference is simply approximated by the error on the reference shear measurement – the true error will only be slightly larger due to the large covariance of the results. The lensing signal around the "BCG-250" is clearly smaller at small projected separations, indicating that a higher fraction of these BCGs are miscentred.
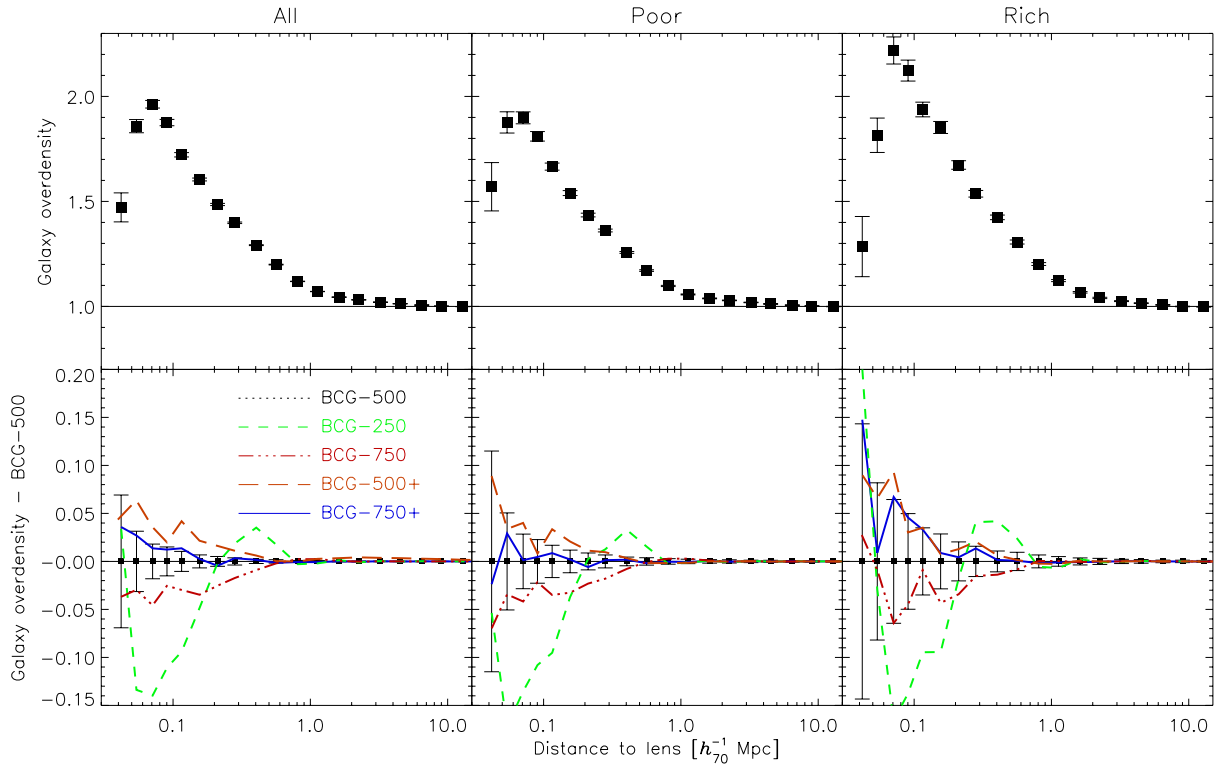

Fig. A.2. Fractional galaxy overdensity around BCG candidates selected with the reference "BCG-500" algorithm (*top panels*). The *bottom panels* show the difference in galaxy overdensity between various BCG selection algorithms and the reference. The *left-hand column* shows the results for all clusters, the middle one for poor clusters only and the *right-hand panel* for rich clusters only. The error on the difference is simply approximated by the error on the reference overdensity measurement – the true error will only be slightly larger due to the large covariance of the results. The "BCG-500+" leads to the highest, most concentrated peak, indicating that these BCGs are closest to the centre of the cluster member distribution.

Both the stacked shear signal of all clusters and the total galaxy overdensity signal around the centre given by the cluster finder, reveal that the clusters have a broad miscentring distribution. In principle, if our cluster halo model is sufficiently flexible, our results should not critically depend on the choice of the cluster centre. However, given that the cluster miscentring distribution is somewhat uncertain, it is better to try to optimise the definition of the centre.

To improve the centring of the clusters we attempt to identify the BCG. For this purpose, we use the catalogue of red-sequence galaxies that were used to identify the cluster. We try a number of different schemes for identifying the BCG and compare the stacked shear signal and galaxy overdensity signal. The most successful prescription will result in a maximal shear and galaxy overdensity signal at small scales; miscentring basically shifts power to larger projected separations.

We start with identifying the brightest red-sequence candidate in the z-band at the cluster redshift within $250\ h_{70}^{-1}$ kpc, $500\ h_{70}^{-1}$ kpc and $750\ h_{70}^{-1}$ kpc from the original cluster centre and adopt it as the candidate BCG. The total lensing signal and galaxy overdensity around these BCG selections are shown in Figs. A.1 and A.2, and are labelled "BCG-250", "BCG-500" and "BCG-750", respectively. The shear signal around the "BCG-250" BCGs is clearly lower at small scales than for the other options. Due to the large errors in the shear signal at small scales, there is no obvious difference in the shear signals between the "BCG-500" and "BCG-750" selections. However, the galaxy overdensity shows that the "BCG-500" BCGs better coincides with the peak of the galaxy distribution, and hence are better centred.

A visual inspection of a number of rich clusters reveals that in some cases, the brightest galaxy is quite far offset from the distribution of cluster members. Hence to improve the "BCG-500" and "BCG-750" selections, we compute the weighted z-band luminosity centre of the cluster using all red-sequence candidates that are located within $500\ h_{70}^{-1}$ kpc from the original cluster's centre. If there exists a second-bright red-sequence candidate member at the cluster redshift that is at most 0.5 mag fainter than the first selected BCG, but that is located at least $100\ h_{70}^{-1}$ kpc closer to the centre of light, we adopt it as the BCG and consequently as the new centre of the cluster. This occurs in 30% and 34% of the "BCG-500" and "BCG-750" selections, respectively. The resulting BCG selections are called "BCG-500+" and "BCG-750+". From Fig. A.2 we observe that the resulting BCG catalogues are indeed better centred. The one with the highest galaxy overdensity is "BCG-500+". We therefore adopt the BCG candidates from this algorithm as the centres of the clusters. The richness estimates of the clusters are recomputed using the new centres.

Appendix B: Impact of cluster redshift scatter

The assigned "red-sequence" redshifts of the clusters have a certain scatter with respect to their actual redshifts. This is caused by intrinsic scatter in the red sequence, uncertainties in the background subtraction, contamination of fore- and/or background galaxies, and noise. Redshift scatter affects the lensing analysis in three ways: it biases the adopted lensing efficiencies, it causes a radial smoothing of the lensing signal, and, if the signal is non-linearly redshift dependent, also an additional smoothing of the signal in that direction.

The lensing efficiency does not linearly scale with lens redshift. Hence if the actual redshifts are scattered compared to the
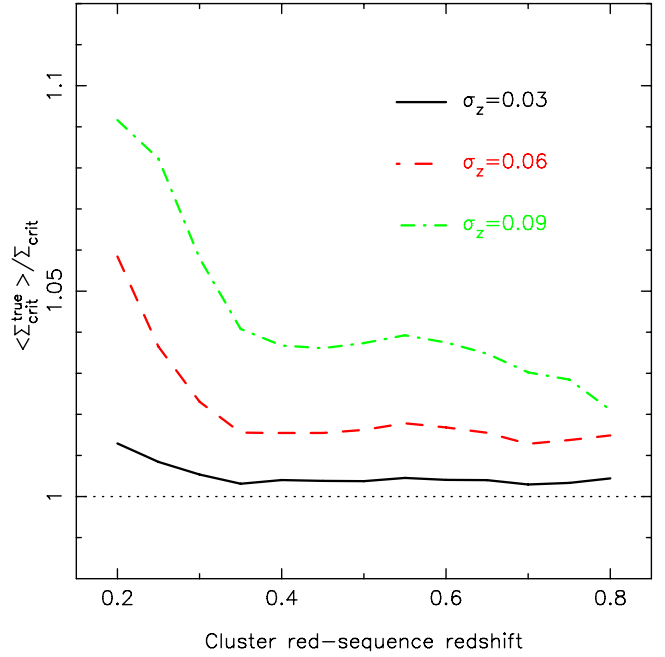

Fig. B.1. Impact of redshift scatter on the lensing efficiencies.

adopted ones, the average over these true lensing efficiencies does not equal the average over the lensing efficiencies of the adopted redshifts. We simulate the impact as follows. For a given cluster "red-sequence" redshift, we assume that the true redshift distribution follows a Gaussian with a certain width (this is identical to assuming that the true distribution of redshifts is flat, and the "red-sequence" redshifts follow a Gaussian distribution around the true value). We draw a large number of true cluster redshifts from this Gaussian, and for each we compute $\Delta\Sigma_{\mathrm{crit}}^{\mathrm{true}}$. If the redshift that is drawn is lower than 0.05 or higher than 1, we disregard it. We average these critical surface mass densities, and compare it to the $\Delta\Sigma_{\mathrm{crit}}$ of the input "red-sequence" redshift. We plot the ratio in Fig. B.1.

The average of the distribution of true $\Delta\Sigma_{\mathrm{crit}}$ is higher than the value of the single "red-sequence" redshift (the one we would use). The size of the bias increases at lower redshifts and for higher values of the scatter. For our cluster sample, the estimated size of redshift scatter is ~0.03. The difference between the mean of the true critical surface mass densities and the one we used is smaller than one per cent over the range of cluster redshifts considered. This is considerably smaller than the statistical error, and we therefore ignore the effect.

Next we investigate the impact of the radial smoothing due to redshift scatter. We only focus on the lensing signal here; the cluster-satellite clustering should be affected in a similar way. We create a typical cluster lensing profile with our halo model for a certain "red-sequence" redshift at fixed physical separations. Next we assume that the true redshift distribution follows a Gaussian with a certain width and draw true redshifts from that. For each redshift, we compute the signal at the same angular separation (but corresponding to different physical separations), and do the radial binning assuming the redshift is the adopted "red-sequence" redshift. Hence we radially smooth the model profile. Then we compute the ratio of the average of this smoothed shear profile and the input profile, and show it in Fig. B.2.

In general, the smoothed profile is larger than the profile at a fixed "red-sequence" redshift. The difference is largest at low redshifts, and increases when the redshift scatter increases.
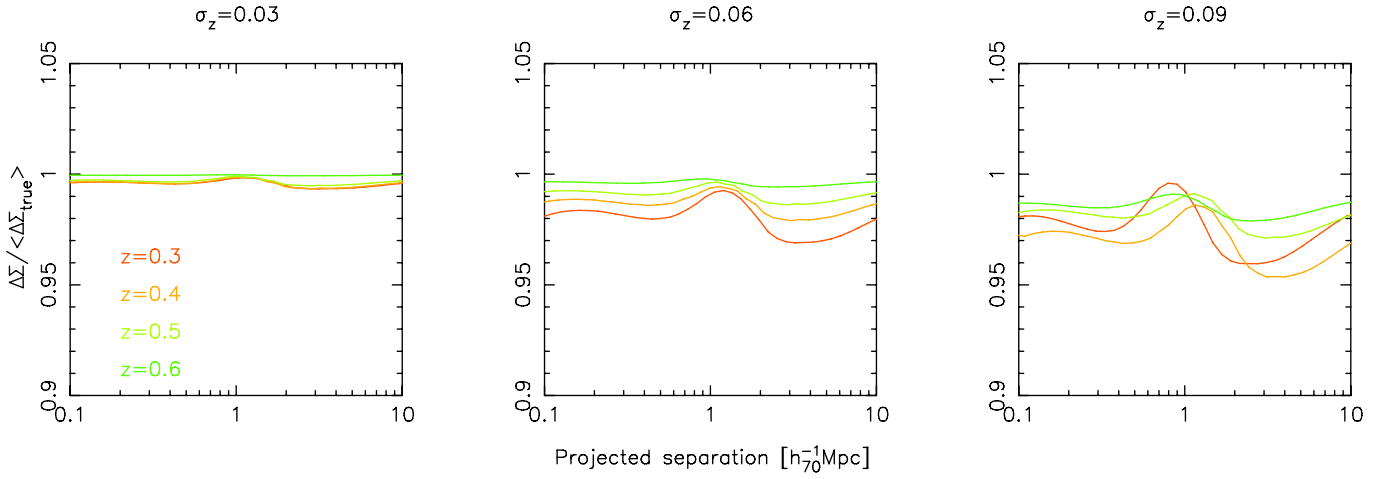
Fig. B.2. Impact of redshift scatter on the radial profile of the model signal.

However, the actual ratio obviously depends on the model profile that we have assumed. Deriving a correction factor in a model independent way is therefore not obvious, and we refrain from doing so. It is also not necessary, given that the impact on the signal is of order a per cent, smaller than our statistical errors.

Finally, there is the effect of additional smoothing in the redshift direction. Given that the scatter is much smaller than the redshift bin size used in this work, this effect is negligible.